(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 872 554 A2

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
21.10.1998 Bulletin 1998/43

(21) Application number: **97120324.5**

(22) Date of filing: **20.11.1997**

(51) Int. Cl.6: **C12N 15/52**, C12P 23/00, A23K 1/16, A23L 1/275, C12N 1/21

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **02.12.1996 EP 96810839**

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)

(72) Inventors:
- **Pasamontes, Luis**
4632 Trimbach (CH)
- **Tsygankov, Yuri,**
State Scient. Cnt. of Russian Fed
Moscow 113545 (RU)

(74) Representative: **Braun, Axel et al**
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Improved fermentative carotenoid production

(57) The present invention is directed to processes for the preparation of canthaxanthin, adonixanthin, astaxanthin, a mixture of adonixanthin and astaxanthin and zeaxanthin by a cell which has been transformed by DNA sequences encoding the respective biosynthetic enzymes of Flavobacterium and the gram negative bacterium E-396. Furthermore the present invention is directed to a food or feed composition comprising one or more of the aforementioned carotenoids.

EP 0 872 554 A2

Description

Over 600 different carotenoids have been described from carotenogenic organisms found among bacteria, yeast, fungi and plants. Currently only two of them, β-carotene and astaxanthin are commercially produced in microorganisms and used in the food and feed industry. β-carotene is obtained from algae and astaxanthin is produced in Pfaffia strains which have been generated by classical mutation. However, fermentation in Pfaffia has the disadvantage of long fermentation cycles and recovery from algae is cumbersome. Therefore it is desiderable to develop production systems which have better industrial applicability, e.g. can be manipulated for increased titers and/or reduced fermentation times. Two such systems using the biosynthetic genes form Erwinia herbicola and Erwinia uredovora have already been described in WO 91/13078 and EP 393 690, respectively. Furthermore, three β-carotene ketolase genes (β-carotene β-4-oxygenase) of the marine bacteria *Agrobacterium aurantiacum* and *Alcaligenes* strain PC-1 (crtW) [Misawa, 1995, Biochem. Biophys. Res. Com. 209, 867-876][Misawa, 1995, J. Bacteriology 177, 6575-6584] and from the green algae *Haematococcus pluvialis* (bkt) [Lotan, 1995, FEBS Letters 364, 125-128][Kajiwara, 1995, Plant Mol. Biol. 29, 343-352] have been cloned. *E. coli* carrying either the carotenogenic genes (crtE, crtB, crtY and crtI) of *E. herbicola* [Hundle, 1994, MGG 245, 406-416] or of *E. uredovora* and complemented with the crtW gene of *A. aurantiacum* [Misawa, 1995] or the bkt gene of *H. pluvialis* [Lotan, 1995][Kajiwara, 1995] resulted in the accumulation of canthaxanthin (β,β-carotene-4,4'-dione), originating from the conversion of β-carotene, via the intermediate echinenone (β,β-carotene-4-one). Introduction of the above mentioned genes (crtW or bkt) into *E. coli* cells harbouring besides the carotenoid biosynthesis genes mentioned above also the crtZ gene of *E. uredovora* [Kajiwara, 1995][Misawa, 1995], resulted in both cases in the accumulation of astaxanthin (3,3'-dihydroxy-β,β-carotene-4,4'-dione). The results obtained with the bkt gene, are in contrast to the observation made by others [Lotan, 1995], who using the same experimental set-up, but introducing the *H. pluvialis* bkt gene in a zeaxanthin (β,β-carotene-3,3'-diol) synthesising *E. coli* host harbouring the carotenoid biosynthesis genes of *E. herbicola*, a close relative of the above mentioned *E. uredovora* strain, did not observe astaxanthin production.

Since there is a continuing need in even more optimized fermentation systems for industrial application it is therefore in the first instance an object of the present invention to provide a process for the preparation of canthaxanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following DNA sequences:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtI) or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous;

e) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E-396 (FERM BP-4283) [$crtW_{E396}$] or a DNA sequence which is substantially homologous;

or a cell which is transformed by a vector comprising DNA sequences specified above under a) to e) and by isolating canthaxanthin from such cells or the culture medium by methods known in the art.

Furthermore it is in the second instance an object of the present invention to provide a process for the preparation of a mixture of adonixanthin and astaxanthin or adonixanthin or astaxanthin alone by a process as mentioned above characterized therein that in addition to the DNA sequences specified under a) to e) the following additional DNA sequence is present:

f) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E-396 (FERM BP-4283) [$crtZ_{E396}$] or a DNA sequence which is substantially homologous;

and the DNA sequence specified under e) is as specified above or the following sequence:

g) a DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crtW) or a DNA

sequence which is substantially homologous;

and isolating the desired mixture of adonixanthin and astaxanthin or adonixanthin or a astaxanthin alone from such cells of the culture medium and separating the desired mixture or carotenoids alone from other carotenoids which might be present by methods known in the art.

Furthermore it is an object of the present invention to provide a process for the preparation of zeaxanthin by a process as claimed in the first instance characterized therein that the DNA sequence as specified under e) is replaced by the DNA sequence as specified under f) in the second instance and by isolating zeaxanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

Furthermore it is an object of the present invention to provide a process for the production of adonixanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following heterologous DNA sequences:

a) a DNA sequence which encodes the GGPP synthase of the microorganism E-396 (FERM BP-4283) [$crtE_{E396}$] or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase the microorganism E-396 (FERM BP-4283) [$crtB_{E396}$] or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of the microorganism E-396 (FERM BP-4283) [$crtI_{E396}$] or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of the microorganism E-396 (FERM BP-4283) [$crtY_{E396}$] or a DNA sequence which is substantially homologous;

e) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E396 (FERM BP-4283) [$crtZ_{E396}$] or a DNA sequence which is substantially homologous; and

f) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E396 (FERM BP-4283) [$crtW_{E396}$] or a DNA sequence which is substantially homologous;

and isolating adonixanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

Further it is an object of the present invention to provide a process as described above characterized therein that the transformed host cell is a prokaryotic host cell, like E. coli, Bacillus or Flavobacter and a process as described above characterized therein that the transformed host cell is a eukaryotice host cell, like yeast or a fungal cell.

Furthermore it is an object of the present invention to provide a DNA sequence comprising one or more DNA sequences selected from the group consisting of:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtI) or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous; and

e) a DNA sequence which encodes the β-carotene hydroxylase of Flavobacterium sp. R1534 (crtZ) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed

by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid or carotenoid mixture is added to food or feed.

Furthermore, a DNA sequence comprising the following DNA sequences is an object of the present invention:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) of a DNA sequence which is substantially homologous; and

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R 1534 (crtI) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of lycopene and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably lycopene or carotenoid mixture, preferably a lycopene comprising mixture is added to food or feed.

Furthermore a DNA sequence comprising the following DNA sequence is also an object of the present invention:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtI) or a DNA sequence which is substantially homologous; and

d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of β-carotene and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably β-carotene or carotenoid mixture, preferably a β-carotene comprising mixture is added to food or feed.

Furthermore a cell which is transformed by the above mentioned DNA sequence comprising subsequences a) to d) or the vector comprising it and a second DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous or a second vector which comprises a DNA

sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous; and a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of echinenone and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably echinenone or carotenoid mixture, preferably an echinenone comprising mixture is added to food or feed.

Furthermore it is an object of the present invention to provide a DNA sequence as mentioned above comprising subsequences a) to d) and a DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous and a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells of the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, especially such a process for the preparation of echinenone or canthaxanthin and a process for the preparation of a food or feed compositing characterized therein that after such a process has been effected the carotenoid, preferably echinenone or canthaxanthin or carotenoid mixture, preferably a echinenone or canthaxanthin containing mixture is added to food or feed.

Furthermore a DNA sequence comprising the following DNA sequences is also an object of the present invention:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtI) or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous; and

e) a DNA sequence which encodes the β-carotene hydroxylase of Flavobacterium sp. R1534 (crtZ) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells of the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeanxanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin or the carotenoid mixture, preferably a zeaxanthin containing mixture is added to food or feed.

Furthermore a DNA sequence as mentioned above comprising subsequences a) to e) and in addition a DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) of a DNA sequence which is substantially homologous is an object of the present invention and to provide a vector comprising such DNA sequence, preferably in form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture or carotenoids from such cells of the culture medium and, in case only one carotenoid is desired

separting it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeaxanthin, adonixanthin or astaxanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin, adonixanthin or astaxanthin or carotenoid mixture, preferably a zeaxanthin, adonixanthin or astaxanthin containing mixture is added to food or feed.

Furthermore a cell which is transformed by the DNA sequence mentioned above comprising subsequences a) to e) or a vector comprising such DNA sequence and a second DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous or a second vector which comprises a DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous is also an object of the present invention and a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium, and in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeanxanthin or adonixanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin or adonixanthin or carotenoid mixture, preferably a zeaxanthin or adonixanthin containing mixture is added to food or feed.

Furthermore it is an object ot the present invention to provide the DNA sequences and vectors as specified before and a process for the preparation of a food or feed composition characterized therein that after a process as specified before has been effected the carotenoid prepared by such process is added to food or feed.

In this context it should be mentioned that the expression "a DNA sequence is substantially homologous" refers with respect to the crtE encoding DNA sequence to a DNA sequence which encodes an amino acid sequence which shows more than 45 %, preferably more than 60 % and more preferably more than 75 % and most preferably more than 90 % identical amino acids when compared to the amino acid sequence of crtE of Flavobacterium sp. 1534 and is the amino acid sequence of a polypeptide which shows the same type of enzymatic activity as the enzyme encoded by crtE of Flavobacterium sp. 1534. In analogy with respect to crtB this means more than 60 %, preferably more than 70 %, more preferably more than 80 % and most preferably more than 90 %; with respect to crtI this means more than 70 %, preferably more than 80 % and most preferably more than 90 %; with respect to crtY this means 55 %, preferably 70 %, more preferably 80 % and most preferably 90 %.

"DNA sequences which are substantially homologous" refer with respect to the $crtW_{E396}$ encoding DNA sequence to a DNA sequence which encodes an amino acid sequence which shows more than 60%, preferably more than 75% and most preferably more than 90% identical amino acids when compared to the amino acid sequence of crtW of the microorganism E 396 (FERM BP-4283) and is the amino acid sequence of a polypeptide which shows the same type of enzymatic activity as the enzyme encoded by crtW of the microorganism E 396. In analogy with respect to $crtZ_{E396}$ this means more than 75%, preferable more than 80% and most preferably more than 90%; with respect to $crtE_{E396}$, $crtB_{E396}$, $crtI_{E396}$, $crtY_{E396}$ and $crtZ_{E396}$ this means more than 80%, preferably more than 90% and most preferably 95%.

DNA sequences in form of genomic DNA, cDNA or synthetic DNA can be prepared as known in the art [see e.g. Sambrook et al., Molecular Cloning, Cold Spring Habor Laboratory Press 1989] or, e.g. as specifically described in Examples 1, 2 or 7. In the context of the present invention it should be noted that all DNA sequences used for the process for production of carotenoids of the present invention encoding crt-gene products can also be prepared as synthetic DNA sequences according to known methods or in analogy to the method specifically described for crtW in Example 7.

The cloning of the DNA-sequences of the present invention from such genomic DNA can than be effected, e.g. by using the well known polymerase chain reaction (PCR) method. The principles of this method are outlined e.g. in PCR Protocols: A guide to Methods and Applications, Academic Press, Inc. (1990). PCR is an in vitro method for producing large amounts of a specific DNA of defined length and sequence from a mixture of different DNA-sequences. Thereby, PCR is based on the enzymatic amplification of the specific DNA fragment of interest which is flanked by two oligonucleotide primers which are specific for this sequence and which hybridize to the opposite strand of the target sequence. The primers are oriented with their 3' ends pointing toward each other. Repeated cycles of heat denaturation of the template, annealing of the primers to their complementary sequences and extension of the annealed primers with a DNA polymerase result in the amplification of the segment between the PCR primers. Since the extension product of each primer can serve as a template for the other, each cycle essentially doubles the amount of the DNA fragment produced in the previous cycle. By utilizing the thermostable Taq DNA polymerase, isolated from the thermophilic bacteria Thermus aquaticus, it has been possible to avoid denaturation of the polymerase which necessitated the addition of enzyme after each heat denaturation step. This development has led to the automation of PCR by a variety of simple temperature-cycling devices. In addition, the specificity of the amplification reaction is increased by allowing the use of higher temperatures for primer annealing and extension. The increased specificity improves the overall yield of amplified products by minimizing the competition by non-target fragments for enzyme and primers. In this way the specific sequence of interest is highly amplified and can be easily separated from the non-specific sequences by methods known in the

art, e.g. by separation on an agarose gel and cloned by methods known in the art using vectors as described e.g. by Holten and Graham in Nucleic Acid Res. 19, 1156 (1991), Kovalic et. al. in Nucleic Acid Res. 19, 4560 (1991), Marchuk et al. in Nucleic Acid Res. 19, 1154 (1991) or Mead et al. in Bio/Technology 9, 657-663 (1991).

The oligonucleotide primers used in the PCR procedure can be prepared as known in the art and described e.g. in Sambrook et al., s.a.

Amplified DNA-sequences can than be used to screen DNA libraries by methods known in the art (Sambrook et al., s.a.) or as specifically described in Examples 1 and 2.

Once complete DNA-sequences of the present invention have been obtained they can be used as a guideline to define new PCR primers for the cloning of substantially homologous DNA sequences from other sources. In addition they and such homologous DNA sequences can be integrated into vectors by methods known in the art and described e.g. in Sambrook et al. (s.a.) to express or overexpress the encoded polypeptide(s) in appropriate host systems. However, a man skilled in the art knows that also the DNA-sequences themselves can be used to transform the suitable host systems of the invention to get overexpression of the encoded polypeptide. Appropriate host systems are for example Bacteria e.g. E. coli, Bacilli as, e.g. Bacillus subtilis or Flavobacter strains. E. coli, which could be used are E. coli K12 strains e.g. M15 [described as DZ 291 by Villarejo et al. in J. Bacteriol. 120, 466-474 (1974)], HB 101 [ATCC No. 33694] or E. coli SG13009 [Gottesman et al., J. Bacteriol. 148, 265-273 (1981)]. Suitable Flavobacter strains can be obtained from any of the culture collections known to the man skilled in the art and listed, e.g. in the journal "Industrial Property" (January 1994, pgs 29-40), like the American Type Culture Collection (ATCC) or the Centralbureau voor Schimmelkultures (CBS) and are, e.g. Flavobacterium sp. R 1534 (ATCC No. 21588, classified as unknown bacterium; or as CBS 519.67) or all Flavobacter strains listed as CBS 517.67 to CBS 521.67 and CBS 523.67 to CBS 525.67, especially R 1533 (which is CBS 523.67 or ATCC 21081, classified as unknown bacterium; see also USP 3,841,967). Further Flavobacter strains are also described in WO 91/03571. Suitable eukaryotic host systems are for example fungi, like Aspergilli e.g. Aspergillus niger [ATCC 9142] or yeasts, like Saccharomyces, e.g. Saccharomyces cerevisiae or Pichia, like pastoris, all available from ATCC.

Suitable vectors which can be used for expression in E. coli are mentioned, e.g. by Sambrook et al. [s.a.] or by Fiers et al. in Procd. 8th Int. Biotechnology Symposium" [Soc. Franc. de Microbiol., Paris (Durand et al., eds.), pp. 680-697 (1988)] or by Bujard et al. in Methods in Enzymology, eds. Wu and Grossmann, Academic Press, Inc. Vol. 155, 416-433 (1987) and Stüber et al. in Immunological Methods, eds. Lefkovits and Pernis, Academic Press, Inc., Vol. IV, 121-152 (1990). Vectors which could be used for expression in Bacilli are known in the art and described, e.g. in EP 405 370, EP 635 572 Procd. Nat. Acad. Sci. USA 81, 439 (1984) by Yansura and Henner, Meth. Enzym. 185, 199-228 (1990) or EP 207 459. Vectors which can be used for expression in fungi are known in the art and described e.g. in EP 420 358 and for yeast in EP 183 070, EP 183 071, EP 248 227, EP 263 311. Vectors which can be used for expression in Flavobacter are known in the art and described in the Examples or, e.g. in Plasmid Technology, edt. by J. Grinsted and P.M. Bennett, Academic Press (1990).

Once such DNA-sequences have been expressed in an appropriate host cell in a suitable medium the carotenoids can be isolated either from the medium in the case they are secreted into the medium or from the host organism and, if necessary separated from other carotenoids if present in case one specific carotenoid is desired by methods known in the art (see e.g. Carotenoids Vol IA: Isolation and Analysis, G. Britton, S. Liaaen-Jensen, H. Pfander; 1995, Birkhäuser Verlag, Basel).

The carotenoids of the present invention can be used in a process for the preparation of food or feeds. A man skilled in the art is familiar with such process. Such compound foods or feeds can further comprise additives or components generally used for such purpose and known in the state of the art.

After the invention has been described in general hereinbefore, the following figures and examples are intended to illustrate details of the invention, without thereby limiting it in any matter.

Figure 1: The biosynthesis pathway for the formation or carotenoids of *Flavobacterium* sp. R1534 is illustrated explaining the enzymatic activities which are encoded by DNA sequences of the present invention.

Figure 2: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized with Probe 46F. The arrow indicated the isolated 2.4 kb XhoI/PstI fragment.

Figure 3: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with ClaI or double digested with ClaI and HindIII. Blots shown in Panel A and B were hybridized to probe A or probe B, respectively (see examples). Both ClaI/HindIII fragments of 1.8 kb and 9.2 kb are indicated.

Figure 4: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized to probe C. The isolated 2.8 kb Sa1I/HindIII fragment is shown by the

arrow.

Figure 5: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized to probe D. The isolated BclI/SphI fragment of approx. 3 kb is shown by the arrow.

Figure 6: Physical map of the organization of the carotenoid biosynthesis cluster in Flavobacterium sp. R1534, deduced from the genomic clones obtained. The location of the probes used for the screening are shown as bars on the respective clones.

Figure 7: Nucleotide sequence of the *Flavobacterium* sp. R1534 carotenoid biosynthesis cluster and its flanking regions. The nucleotide sequence is numbered from the first nucleotide shown (see BamHI site of Fig. 6). The deduced amino acid sequence of the ORF's (orf-5, orf-1, crtE, crtB, crtI, crtY, crtZ and orf-16) are shown with the single-letter amino acid code. Arrow (-->) indicate the direction of the transcription; asterisks, stop codons.

Figure 8: Protein sequence of the GGPP synthase (crtE) of *Flavobacterium* sp. R1534 with a MW of 31331 Da.

Figure 9: Protein sequence of the prephytoene synthetase (crtB) of *Flavobacterium* sp. R1534 with a MW of 32615 Da.

Figure 10: Protein sequence of the phytoene desaturase (crtI) of *Flavobacterium* sp. R1534 with a MW of 54411 Da.

Figure 11: Protein sequence of the lycopene cyclase (crtY) of *Flavobacterium* sp. R1534 with a MW of 42368 Da.

Figure 12: Protein sequence of the β-carotene hydroxylase (crtZ) of *Flavobacterium* sp. R1534 with a MW of 19282 Da.

Figure 13: Recombinant plasmids containing deletions of the *Flavobacterium* sp. R1534 carotenoid biosynthesis gene cluster.

Figure 14: Primers used for PCR reactions. The underlined sequence is the recognition site of the indicated restriction enzyme. Small caps indicate nucleotides introduced by mutagenesis. Boxes show the artificial RBS which is recognized in B. subtilis. Small caps in bold show the location of the original adenine creating the translation start site (ATG) of the following gene (see original operon). All the ATG's of the original Flavobacter carotenoid biosynthetic genes had to be destroyed to not interfere with the rebuild transcription start site. Arrows indicate start and ends of the indicated Flavobacterium R1534 WT carotenoid genes.

Figure 15: Linkers used for the different constructions. The underlined sequence is the recognition site of the indicated restriction enzyme. Small caps indicate nucleotides introduced by synthetic primers. Boxes show the artificial RBS which is recognized in B. subtilis. Arrow indicate start and ends of the indicated Flavobacterium carotenoid genes.

Figure 16: Costruction of plasmids pBIIKS(+)-clone59-2, pLyco and pZea4.

Figure 17: Construction of plasmid p602CAR.

Figure 18: Construction of plasmids pBIIKS(+)-CARVEG-E and p602 CARVEG-E.

Figure 19: Construction of plasmids pHP13-2CARZYIB-EINV and pHP13-2PN25ZYIB-EINV.

Figure 20: Construction of plasmid pXI12-ZYIB-EINVMUTRBS2C.

Figure 21: Norhern blot analysis of B. subtilis strain BS1012::ZYIB-EINV4. Panel A: Schematic representation of a reciprocal integration of plasmid pXI12-ZYIB-EINV4 into the levan-sucrase gene of B-subtilis. Panel B: Northern blot obtained with probe A (PCR fragment which was obtained with CAR 51 and CAR 76 and

hybridizes to the 3' end of crtZ and the 5' end or crtY). Panel C: Northern blot obtained with probe B (BamHI-XhoI fragment isolated from plasmid pBIIKS(+)-crtE/2 and hybridizing to the 5' part of the crtE gene).

Figure 22: Schematic representation of the integration sites of three transformed Bacillus subtilis strains: BS1012::SFCO, BS1012::SFCOCAT1 and BA1012::SFCONEO1. Amplification of the synthetic Flavobacterium carotenoid operon (SFCO) can only be obtained in those strains having amplifiable structures. Probe A was used to determine the copy number of the integrated SFCO. Erythromycine resistance gene (ermAM), chloramphenicol resistance gene (cat), neomycine resistance gene (neo), terminator of the cryT gene of B. subtilis (cryT), levan-sucrase gene (sac-B 5' and sac-B 3'), plasmid sequences of pXI12 (pXI12), promoter originating from site I of the veg promoter complex (PvegI).

Figure 23: Construction of plasmids pXI12-ZYIB-EINV4MUTRBS2CNEO and pXI12-ZYIB-EINV4MUTRBS2CCAT.

Figure 24: Complete nucleotide sequence of plasmid pZea4.

Figure 25: Synthetic crtW gene of Alcaligenes PC-1. The translated protein sequence is shown above the double stranded DNA sequence. The twelve oligonucleotides (crtW1-crtW12) used for the PCR synthesis are underlined.

Figure 26: Construction of plasmid pBIIKS-crtEBIYZW. The HindIII-Pm1I fragment of pALTER-Ex2-crtW, carrying the synthetic crtW gene, was cloned into the HindIII and MluI (blunt) sites. PvegI and Ptac are the promoters used for the transcription of the two opera. The ColE1 replication origin of this plasmid is compatible with the p15A origin present in the pALTER-Ex2 constructs.

Figure 27: Relevant inserts of all plasmids constructed in Example 7. Disrupted genes are shown by //. Restriction sites: S=SacI, X=XbaI, H=HindIII, N=NsiI, Hp=HpaI, Nd=NdeI.

Figure 28: Reaction products (carotenoids) obtained from β-carotene by the process of the present invention.

Example 1

Materials and general methods used

Bacterial strains and plasmids: *Flavobacterium sp.* R1534 WT (ATCC 21588) was the DNA source for the genes cloned. Partial genomic libraries of *Flavobacterium sp.* R1534 WT DNA were constructed into the pBluescriptII+(KS) or (SK) vector (Stratagene, La Jolla, USA) and transformed into *E. coli* XL-1 blue (Stratagene) or JM109.

Media and growth conditions: Transformed *E. coli* were grown in Luria broth (LB) at 37° C with 100mg Ampicillin (Amp)/ml for selection. *Flavobacterium sp.* R1534 WT was grown at 27° C in medium containing 1% glucose, 1% tryptone (Difco Laboratories), 1% yeast extract (Difco), 0.5% $MgSO_4$ $7H_2O$ and 3% NaCl.

Colony screening: Screening of the *E. coli* transformants was done by PCR basically according to the method described by Zon et al. [Zon et al., BioTechniques 7, 696-698 (1989)] using the following primers:

Primer #7: 5'-CCTGGATGACGTGCTGGAATATTCC-3'
Primer #8: 5'-CAAGGCCCAGATCGCAGGCG-3'

Genomic DNA: A 50 ml overnight culture of *Flavobacterium sp. R1534* was centrifuged at 10,000 g for 10 minutes. The pellet was washed briefly with 10 ml of lysis buffer (50 mM EDTA, 0.1M NaCl pH7.5), resuspended in 4 ml of the same buffer sumplemented with 10 mg of lysozyme and incubated at 37°C for 15 minutes. After addition of 0.3 ml of N-Lauroyl sarcosine (20%) the incubation at 37°C was continued for another 15 minutes before the extraction of the DNA with phenol, phenol/chloroform and chloroform. The DNA was ethanol precipitated at room temperature for 20 minutes in the presence of 0.3 M sodium acetate (pH 5.2), followed by centrifugation at 10,000 g for 15 minutes. The pellet was rinsed with 70% ethanol, dried and resuspended in 1 ml of TE (10 mM Tris, 1mM EDTA, pH 8.0).

All genomic DNA used in the southern blot analysis and cloning experiments was dialysed against H_2O for 48 hours, using collodium bags (Sartorius, Germany), ethanol precipitated in the presence of 0.3 M sodium acetate and resuspended in H_2O.

Probe labelling: DNA probes were labeled with (a - ^{32}P) dGTP (Amersham) by random-priming according to [Sambrook et al., s.a.].

Probes used to screen the mini-libraries: **Probe 46F** is a 119 bp fragment obtained by PCR using primer #7 and #8 and *Flavobacterium sp.* R1534 genomic DNA as template. This probe was proposed to be a fragment of the *Flavobacterium sp.* R1534 phytoene synthase (crtB) gene, since it shows significant homology to the phytoene synthase genes from other species (e.g. *E. uredovora, E. herbicola*). **Probe A** is a BstXI - PstI fragment of 184 bp originating from the right arm of the insert of clone 85. **Probe B** is a 397 bp XhoI - NotI fragment obtained from the left end of the insert of clone 85. **Probe C** is a 536 bp BglII - PstI fragment from the right end of the insert of clone 85. **Probe D** is a 376 bp KpnI - BstYI fragment isolated from the insert of clone 59. The localization of the individual probes is shown in figure 6.

Oligonucleotide synthesis: The oligonucleotides used for PCR reactions or for sequencing were synthesized with an Applied Biosystems 392 DNA synthesizer.

Southern blot analysis: For hybridization experiments *Flavobacterium sp. R1534* genomic DNA (3 mg) was digested with the appropiate restriction enzymes and electrophoresed on a 0.75% agarose gel. The transfer to Zeta-Probe blotting membranes (BIO-RAD), was done as described [Sourthern, E.M., J. Mol. Biol. 98, 503 (1975)]. Prehybridization and hybridization was in 7%SDS, 1% BSA (fraction V; Boehringer), 0.5M Na_2HPO_4, pH 7.2 at 65°C. After hybridization the membranes were washed twice for 5 minutes in 2x SSC, 1% SDS at room temperature and twice for 15 minutes in 0.1% SSC, 0.1% SDS at 65° C.

DNA sequence analysis: The sequence was determined by the dideoxy chain termination technique [Sanger et al., Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977)] using the Sequenase Kit (United States Biochemical). Both strands were completely sequenced and the sequence analyzed using the GCG sequence analysis software package (Version 8.0) by Genetics Computer, Inc. [Devereux et al., Nucleic Acids. Res. 12, 387-395 (1984)].

Analysis of carotenoids: *E. coli* XL-1 or JM109 cells (200 - 400 ml) carrying different plasmid constructs were grown for the times indicated in the text, usually 24 to 60 hours, in LB suplemented with 100mg Ampicillin/ml, in shake flasks at 37° C and 220 rpm.

The carotenoids present in the microorganisms were extracted with an adequate volume of acetone using a rotation homogenizer (Polytron, Kinematica AG, CH-Luzern). The homogenate was the filtered through the sintered glass of a suction filter into a round bottom flask. The filtrate was evaporated by means of a rotation evaporator at 50° C using a water-jet vacuum. For the zeaxanthin detection the residue was dissolved in n-hexane/acetone (86:14) before analysis with a normalphase HPLC as described in [Weber, S. in Analytical Methods for Vitamins and Carotenoids in Feed, Keller, H.E. , Editor, 83-85 (1988)]. For the detection of β-carotene and lycopene the evaporated extract was dissolved in n-hexane/acetone (99:1) and analysed by HPLC as described in [Hengartner et al., Helv. Chim. Acta 75, 1848-1865 (1992)].

Example 2

Cloning of the *Flavobacterium sp.* R1534 carotenoid biosynthetic genes.

To identify and isolate DNA fragments carrying the genes of the carotenoid biosynthesis pathway, we used the DNA fragment 46F (see methods) to probe a Southern Blot carrying chromosomal DNA of *Flavobacterium sp.* R1534 digested with different restriction enzymes Fig. 2. The 2.4 kb XhoI/PstI fragment hybridizing to the probe seemed the most appropiate one to start with. Genomic *Flavobacterium sp.* R1534 DNA was digested with XhoI/PstI and run on a 1% agarose gel. According to a comigrating DNA marker, the region of about 2.4 kb was cut out of the gel and the DNA isolated. A XhoI/PstI mini library of *Flavobacterium sp.* R1534 genomic DNA was constructed into XhoI - PstI sites of pBluescriptIISK(+). One hundred *E. coli* XL1 transformants were subsequentely screened by PCR with primer #7 and primer # 8, the same primers previously used to obtain the 119 bp fragment (46F). One positive transformant, named clone 85, was found. Sequencing of the insert revealed sequences not only homologous to the phytoene synthase (crtB) but also to the phytoene desaturase (crtI) of both *Erwinia* species herbicola and uredovora. Left and right hand genomic sequences of clone 85 were obtained by the same approach using probe A and probe B. *Flavobacterium sp.* R1534 genomic DNA was double digested with ClaI and Hind III and subjected to Southern analysis with probe A and probe B. With probe A a ClaI/HindIII fragment of aprox. 1.8 kb was identified (Fig. 3A), isolated and subcloned into the ClaI/HindIII sites of pBluescriptIIKS (+). Screening of the *E. coli* XL1 transformants with probe A gave 6 positive clones. The insert of one of these positives, clone 43-3, was sequenced and showed homology to the N-terminus of crtI genes and to the C-terminus of crtY genes of both *Erwinia* species mentioned above. With probe B an approx. 9.2 kb ClaI/HindIII fragment was detected (Fig. 3B), isolated and subcloned into pBluescriptIIKS (+).

A screening of the transformants gave one positive, clone 51. Sequencing of the 5' and 3' of the insert, revealed that only the region close to the HindIII site showed relevant homology to genes of the carotenoid biosynthesis of the *Erwinia* species mentioned above (e.g. crtB gene and crtE gene). The sequence around the ClaI site showed no homology to known genes of the carotenoid biosynthesis pathway. Based on this information and to facilitate further sequencing and construction work, the 4.2 kb BamHI/HindIII fragment of clone 51 was subcloned into the respective sites of pBluescriptIIKS(+) resulting in clone 2. Sequencing of the insert of this clone confirmed the presence of genes homol-

ogous to *Erwinia sp.* crtB and crtE genes. These genes were located within 1.8 kb from the HindIII site. The remaining 2.4 kb of this insert had no homology to known carotenoid biosynthesis genes.

Additional genomic sequences downstream of the ClaI site were detected using probe C to hybridize to *Flavobacterium sp.* R1534 genomic DNA digested with different restriction enzymes (see figure 4).

A SalI/HindIII fragment of 2.8 kb identified by Southern analysis was isolated and subcloned into the HindIII/XhoI sites of pBluescriptIIKS (+). Screening of the *E. coli* XL1 transformants with probe A gave one positive clone named clone 59. The insert of this clone confirmed the sequence of clone 43-3 and contained in addition sequences homologous to the N-terminus of the crtY gene from other known lycopene cyclases. To obtain the putative missing crtZ gene a Sau3AI partial digestion library of *Flavobacterium sp. R1534* was constructed into the BamHI site of pBluescriptIIKS(+). Screening of this library with probe D gave several positive clones. One transformant designated, clone 6a, had an insert of 4.9 kb. Sequencing of the insert revealed besides the already known sequences coding for crtB, crtI and crtY also the missing crtZ gene. Clone 7g was isolated from a mini library carrying BclI/SphI fragments of R1534 (Fig. 5) and screened with probe D. The insert size of clone 7g is approx. 3 kb.

The six independent inserts of the clones described above covering approx. 14 kb of the *Flavobacterium sp.* R1534 genome are compiled in Figure 6.

The determined sequence spanning from the BamHI site (position 1) to base pair 8625 is shown figure 7.

Putative protein coding regions of the cloned R1534 sequence.

Computer analysis using the CodonPreference program of the GCG package, which recognizes protein coding regions by virtue of the similarity of their codon usage to a given codon frequency table, revealed eight open reading frames (ORFs) encoding putative proteins: a partial ORF from 1 to 1165 (ORF-5) coding for a polypeptide larger than 41382 Da; an ORF coding for a polypeptide with a molecular weight of 40081 Da from 1180 to 2352 (ORF-1); an ORF coding for a polypeptide with a molecular weight of 31331 Da from 2521 to 3405 (crtE); an ORF coding for a polypeptide with a molecular weight of 32615 Da from 4316 to 3408 (crtB); an ORF coding for a polypeptide with a molecular weight of 54411 Da from 5797 to 4316 (crtI); an ORF coding for a polypeptide with a molecular weight of 42368 Da from 6942 to 5797 (crtY); an ORF coding for a polypeptide with a molecular weight of 19282 Da from 7448 to 6942 (crtZ); and an ORF coding for a polypeptide with a molecular weight of 19368 Da from 8315 to 7770 (ORF-16); ORF-1 and crtE have the opposite transcriptional orientation from the others (Fig. 6). The translation start sites of the ORFs crtI, crtY and crtZ could clearly be determined based on the appropiately located sequences homologous to the Shine/Delgano (S/D) [Shine and Dalgarno, Proc. Natl. Acad. Sci. USA 71, 1342-1346 (1974)] consensus sequence AGG--6-9N--ATG (Fig. 10) and the homology to the N-terminal sequences of the respective enzymes of *E. herbicola* and *E. uredovora*. The translation of the ORF crtB could potentially start from three closely spaced codons ATG (4316), ATG (4241) and ATG (4211). The first one, although not having the best S/D sequence of the three, gives a translation product with the highest homology to the N-terminus of the *E. herbicola* and *E. uredovora* crtB protein, and is therefore the most likely translation start site. The translation of ORF crtE could potentially start from five different start codons found within 150 bp : ATG (2389), ATG (2446), ATG (2473), ATG (2497) and ATG (2521). We believe that based on the following observations, the ATG (2521) is the most likely transcription start site of crtE: this ATG start codon is preceeded by the best consensus S/D sequence of all five putative start sites mentioned; and the putative N-terminal amino acid sequence of the protein encoded has the highest homology to the N-terminus of the crtE enzymes of *E. herbicola* and *E. uredovora*;

Characteristics of the crt translational initiation sites and gene products.

The translational start sites of the five carotenoid biosynthesis genes are shown below and the possible ribosome binding sites are underlined. The genes crtZ, crtY, crtI and crtB are grouped so tightly that the TGA stop codon of the anterior gene overlaps the ATG of the following gene. Only three of the five genes (crtI, crtY and crtZ) fit with the consensus for optimal S/D sequences. The boxed TGA sequence shows the stop condon of the anterior gene.

-10	+1	
ACGAAGGCACCGATGACGCCCA		crtE
CGGACCTGGCCGTCGCATGACCGATC		crtB
CGGATCGCAATACATGAGCCATG		crtY
CTGCAGGAGAGAGCATGAGTTCCG		crtI
GCAAGGGGCCGGCATGAGCACTT		crtZ

Amino acid sequences of individual crt genes of *Flavobacterium sp.* R1534.

All five ORFs of *Flavobacterium sp.* R1534 having homology to known carotenoid biosynthesis genes of other species are clustered in approx. 5.2 kb of the sequence (Fig. 7).

GGDP synthase (crtE)

The amino acid (aa) sequence of the geranylgeranyl pyrophosphate synthase (crtE gene product) consists of 295 aa and is shown in figure 8. This enzyme condenses farnesyl pyrophosphate and isopentenyl pyrophosphate in a 1' - 4.

Phytoene synthase (crtB)

This enzyme catalyzes two enzymatic steps. First it condenses in a head to head reaction two geranylgeranyl pyrophosphates (C20) to the C40 carotenoid prephytoene. Second it rearanges the cyclopropylring of prephytoene to phytoene. The 303 aa encoded by the crtB gene of *Flavobacterium sp.* R1534 is shown in figure 9.

Phytoene desaturase (crtI)

The phytoene desaturase of *Flavobacterium sp.* R1534 consisting of 494 aa, shown in figure 10, performs like the crtI enzyme of *E. herbicola* and *E. uredovora*, four desaturation steps, converting the non-coloured carotenoid phytoene to the red coloured lycopene. ***Lycopene cyclase (crtY)***

The crtY gene product of *Flavobacterium sp.* R1534 is sufficient to introduce the b-ionone rings at both sides of lycopene to obtain β-carotene. The lycopene cyclase of *Flavobacterium sp.* R1534 consists of 382 aa (Fig. 11). ***β-carotene hydroxylase (crtZ)***

The gene product of crtZ consisting of 169 aa (Fig. 12) and hydroxylates β-carotene to the xanthophyll zeaxanthin.

Putative enzymatic functions of the ORF's (orf-1, orf-5 and orf-16)

The orf-1 has at the aa level over 40% identity to acetoacetyl-CoA thiolases of different organisms (e.g. Candida tropicalis, human, rat). This gene is therefore most likely a putative acetoacetyl-CoA thiolase (acetyl-CoA acetyltransferase), which condenses two molecules of acetyl-CoA to Acetoacetyl-CoA. Condensation of acetoacetyl-CoA with a third acetyl-CoA by the HMG-CoA synthase forms β-hydroxy-β-methylglutaryl-CoA (HMG-CoA). This compound is part of the mevalonate pathway which produces besides sterols also numerous kinds of isoprenoids with diverse cellular functions. In bacteria and plants, the isoprenoid pathway is also able to synthesize some unique products like carotenoids, growth regulators (e.g. in plants gibberellins and abcissic acid) and sencodary metabolites like phytoalexins [Riou et al., Gene 148, 293-297 (1994)].

The orf-5 has a low homology of approx. 30%, to the amino acid sequence of polyketide synthases from different streptomyces (e.g. S. violaceoruber, S. cinnamonensis). These antibiotic synthesizing enzymes (polyketide synthases), have been classified into two groups. Type-I polyketide synthases are large multifunctional proteins, whereas type-II

polyketide synthases are multiprotein complexes composed of several individual proteins involved in the subreactions of the polyketide synthesis [Bibb, et al. Gene 142, 31-39 (1994)].

The putative protein encoded by the orf-16 has at the aa level an identity of 42% when compared to the soluble hydrogenase subunit of Anabaena cylindrica.

Functional assignment of the ORF 's (crtE, crtB, crtI, crtY and crtZ) to enzymatic activities of the carotenoid biosynthesis pathway.

The biochemical assignment of the gene products of the different ORF's were revealed by analyzing carotenoid accumulation in *E. coli* host strains that were transformed with deleted variants of the *Flavobacterium sp.* gene cluster and thus expressed not all of the crt genes (Fig. 13).

Three different plasmid were constructed: pLyco, p59-2 and pZea4. Plasmid p59-2 was obtained by subcloning the HindIII/BamHI fragment of clone 2 into the HindIII/BamHI sites of clone 59. p59-2 carries the ORF's of the crtE, crtB, crtI and crtY gene and should lead to the production of β-carotene. pLyco was obtained by deleting the KpnI/KpnI fragment, coding for approx. one half (N-terminus) of the crtY gene, from the p59-2 plasmid. *E. coli* cells transformed with pLyco, and therefore having a truncated non-functional crtY gene, should produce lycopene, the precursor of β-carotene. pZea4 was constructed by ligation of the AscI-SpeI fragment of p59-2, containing the crtE, crtB, crtI and most of the crtY gene with the AscI/XbaI fragment of clone 6a, containing the sequences to complete the crtY gene and the crtZ gene. pZea4 [for complete sequence see Fig. 24; nucleotides 1 to 683 result from pBluescriptIIKS(+), nucleotides 684 to 8961 from Flavobacterium R1534 WT genome, nucleotides 8962 to 11233 from pBluescriptIIKS(+)] has therefore all five ORF's of the zeaxanthin biosynthesis pathway. Plasmid pZea4 has been deposited on May 25, 1995 at the DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany) under accession No. DSM 10012. *E. coli* cells transformed with this latter plasmid should therefore produce zeaxanthin. For the detection of the carotenoid produced, transformants were grown for 48 hours in shake flasks and then subjected to carotenoid analysis as described in the methods section. Figure 13 summarizes the different inserts of the plasmids described above, and the main carotenoid detected in the cells.

As expected the pLyco carrying *E. coli* cells produced lycopene, those carrying p59-2 produced β-carotene (all-E,9-Z,13-Z) and the cells having the pZea4 construct produced zeaxanthin. This confirms that all the necessary genes of *Flavobacterium sp.* R1534 for the synthesis of zeaxanthin or their precursors (phytoene, lycopene and β-carotene) were cloned.

Example 3

Materials and methods used for expression of carotenoid synthesizing enzymes

Bacterial strains and plasmids: The vectors pBluescriptIIKS (+) or (-) (Stratagene, La Jolla, USA) and pUC18 [Vieira and Messing, Gene 19, 259-268 (1982); Norrander et al., Gene 26, 101-106 (1983)] were used for cloning in different *E. coli* strains, like XL-1 blue (Stratagene), TG1 or JM109. In all *B. subtilis* transformations, strain 1012 was used. Plasmids pHP13 [Haima et al., Mol. Gen. Genet. 209, 335-342 (1987)] and p602/22 [LeGrice, S.F.J. in Gene Expression Technology, Goeddel, D.V., Editor, 201-214 (1990)] are Gram (+)/(-) shuttle vectors able to replicate in *B. subtilis* and *E. coli* cells. Plasmid p205 contains the vegI promoter cloned into the SmaI site of pUC18. Plasmid pXI12 is an integration vector for the constitutive expression of genes in *B. subtilis* [Haiker et al., in 7th Int. Symposium on the Genetics of Industrial Microorganisms, June 26-July 1, 1994. Mongreal, Quebec, Canada (1994)]. Plasmid pBEST501 [Itaya et al., Nucleic Acids Res. 17 (11), 4410 (1989)] contains the neomycin resistance gene cassette originating from the plasmid pUB110 (GenBank entry: M19465) of *S. aureus* [McKenzie et al., Plasmid 15, 93-103 (1986); McKenzie et al., Plasmid 17, 83-84 (1987)]. This neomycin gene has been shown to work as a selection marker when present in a single copy in the genome of *B. subtilis*. Plasmid pC194 (ATCC 37034)(GenBank entry: L08860) originates from *S. aureus* [Horinouchi and Weisblaum, J. Bacteriol. 150, 815-825 (1982)] and contains the chloramphenicol acetyltransferase gene.

Media and growth conditions: *E. coli* were grown in Luria broth (LB) at 37° C with 100mg Ampicillin (Amp)/ml for selection. *B. subtilis* cells were grown m VY-medium supplemented with either erythromycin (1 mg/ml), neomycin (5-180 mg/ml) or chloramphenicol (10-80 mg/ml).

Transformation: *E. coli* transformations were done by electroporation using the Gene-pulser device of BIO-RAD (Hercules, CA, USA) with the following parameters (200 W, 250 mFD, 2.5V). *B.subtilis* transformations were done basically according to the standard procedure method 2.8 described by [Cutting and Vander Horn in Molecular Biological Methods for Bacillus, Harwood, C.R. and Cutting, S.M., Editor, John Wiley & Sons: Chichester, England. 61-74 (1990)].

Colony screening: Bacterial colony screening was done as described by [Zon et al., s.a.].

Oligonucleotide synthesis: The oligonucleotides used for PCR reactions or for sequencing were synthesized with

an Applied Biosystems 392 DNA synthesizer.

PCR reactions: The PCR reactions were performed using either the **UlTma DNA polymerase** (Perkin Elmer Cetus) or the **Pfu Vent polymerase** (New England Biolabs) according to the manufacturers instructions. A typical 50 ml PCR reaction contained: 100ng of template DNA, 10 pM of each of the primers, all four dNTP's (final conc. 300 mM), $MgCl_2$ (when UlTma polymerase was used; final conc. 2 mM), 1x UlTma reaction buffer or 1x Pfu buffer (supplied by the manufacturer). All components of the reaction with the exception of the DNA polymerase were incubated at 95°C for 2 min. followed by the cycles indicated in the respective section (see below). In all reactions a hot start was made, by adding the polymerase in the first round of the cycle during the 72°C elongation step. At the end of the PCR reaction an aliquot was analysed on 1% agarose gel, before extracting once with phenol/chloroform. The amplified fragment in the aqueous phase was precipitated with 1/10 of a 3M NaAcetate solution and two volumes of Ethanol. After centrifugation for 5 min. at 12000 rpm, the pellet was resuspended in an adequate volume of H_2O, typically 40 ml, before digestion with the indicated restriction enzymes was performed. Afier the digestion the mixture was separated on a 1% low melting point agarose. The PCR product of the expected size were excised from the agarose and purified using the glass beads method (GENECLEAN KIT, Bio 101, Vista CA, USA) when the fragments were above 400 bp or directly spun out of the gel when the fragments were shorter than 400 bp as described by [Heery et al., TIBS 6 (6), 173 (1990)].

Oligos used for gene amplification and site directed mutagenesis:

All PCR reactions performed to allow the construction of the different plasmids are described below. All the primers used are summarized in figure 14.

Primers #100 and #101 were used in a PCR reaction to amplify the complete crtE gene having a SpeI restriction site and an artificial ribosomal binding site (RBS) upstream of the transcription start site of this gene. At the 3' end of the amplified fragment, two unique restriction sites were introduced, an AvrII and a SmaI site, to facilitate the further cloning steps. The PCR reaction was done with UlTma polymerase using the following conditions for the amplification: 5 cycles with the profile: 95°C, 1 min./ 60°C, 45 sec./ 72°C, 1 min. and 20 cycles with the profile: 95°C, 1 min./ 72°C, 1 min.. Plasmid pBIIKS(+)-clone2 served as template DNA. The final PCR product was digested with SpeI and SmaI and isolated using the GENECLEAN KIT. The size of the fragment was approx. 910 bp.

Primers **#104** and **#105** were used in a PCR reaction to amplify the crtZ gene from the translation start till the SalI restriction site, located in the coding sequence of this gene. At the 5' end of the crtZ gene an EcoRI, a synthetic RBS and a NdeI site was introduced. The PCR conditions were as described above. Plasmid pBIIKS(+)-clone 6a served as template DNA and the final PCR product was digested with EcoRI and SalI. Isolation of the fragment of approx. 480 bp was done with the GENECLEAN KIT.

Primers **MUT1** and **MUT5** were used to amplify the complete crtY gene. At the 5' end, the last 23 nucleotides of the crtZ gene including the SalI site are present, followed by an artificial RBS preceding the translation start site of the crtY gene. The artificial RBS created includes a PmlI restriction site. The 3' end of the amplified fragment contains 22 nucleotides of the crtI gene, preceded by an newly created artifial RBS which contains a MunI restriction site. The conditions used for the PCR reaction were as described above using the following cycling profile: 5 rounds of 95°C, 45 sec./ 60°C, 45 sec./ 72°C, 75 sec. followed by 22 cycles with the profile: 95°C, 45 sec./ 66°C, 45 sec./ 72°C, 75 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. The PCR product of 1225 bp was made blunt and cloned into the SmaI site of pUC18, using the Sure-Clone Kit (Pharmacia) according to the manufacturer.

Primers **MUT2** and **MUT6** were used to amplify the complete crtI gene. At the 5' the last 23 nucleotides of the crtY gene are present, followed by an artificial RBS which precedes the translation start site of the crtI gene. The new RBS created, includes a MunI restriction site. The 3' end of the amplified fragment contains the artificial RBS upstream of the crtB gene including a BamHI restriction site. The conditions used for the PCR reaction were basically as described above including the following cycling profile: 5 rounds of 95°C, 30 sec./ 60°C, 30 sec./ 72°C, 75 sec., followed by 25 cycles with the profile: 95°C, 30 sec./ 66°C, 30 sec./ 72°C, 75 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. For the further cloning steps the PCR product of 1541 bp was digested with MunI and BamHI.

Primers **MUT3** and **CAR17** were used to amplify the N-terminus of the crtB gene. At the 5' the last 28 nucleotides of the crtI gene are present followed by an artificial RBS, preceding the translation start site of the crtB gene. This new created RBS, includes a BamHI restriction site. The amplified fragment, named PCR-F contains also the HindIII restriction site located at the N-terminus of the crtB gene. The conditions used for the PCR reaction were as described elsewhere in the text, including the following cycling profile: 5 rounds of 95°C, 30 sec./ 58°C, 30 sec./ 72°C, 20 sec. followed by 25 cycles with the profile: 95°C, 30 sec./ 60°C, 30 sec./ 72°C, 20 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. The PCR product of approx. 160 bp was digested with BamHI and HindIII.

Olig s used to amplify the chloramphenicol resistance gene (cat):

Primers **CAT3** and **CAT4** were used to amplify the chloramphenicol resistance gene of pC194 (ATCC 37034) [Hori-

nouchi and Weisblum, s.a.] a R-plasmid found in *S. aureus*. The conditions used for the PCR reaction were as described previously including the following cycling profile: 5 rounds of 95°C, 60 sec./ 50°C, 60 sec./ 72°C, 2 min. followed by 20 cycles with the profile: 95°C, 60 sec./ 60°C, 60 sec./ 72°C, 2 min.. Plasmid pC198 served as template for the Pfu Vent polymerase. The PCR product of approx. 1050 bp was digested with EcoRI and AatII.

Oligos used to generate linkers: Linkers were obtained by adding 90 ng of each of the two corresponding primers into an Eppendorf tube. The mixture was dried in a speed vac and the pellet resuspended in 1x Ligation buffer (Boehringer, Mannheim, Germany). The solution was incubated at 50°C for 3 min. before cooling down to RT, to allow the primers to hybridize properly. The linker were now ready to be ligated into the appropriate sites. All the oligos used to generate linkers are shown in figure 15.

Primers **CS1** and **CS2** were used to form a linker containing the following restrictions sites HindIII, AflII, ScaI, XbaI, PmeI and EcoRI.

Primers **MUT7** and **MUT8** were used to form a linker containing the restriction sites SalI, AvrII, PmlI, MluI, MunI, BamHI, SphI and HindIII.

Primers **MUT9** and **MUT10** were used to introduce an artificial RBS upstream of crtY.

Primers **MUT11** and **MUT12** were used to introduce an artificial RBS upstream of crtE.

Isolation of RNA: Total RNA was prepared from log phase growing *B. subtilis* according to the method described by [Maes and Messens, Nucleic Acids Res. 20 (16), 4374 (1992)].

Northern Blot analysis: For hybridization experiments up to 30 mg of *B. subtilis* RNA was electrophoreses on a 1% agarose gel made up in 1x MOPS and 0.66 M formaldehyde. Transfer to Zeta-Probe blotting membranes (BIO-RAD), UV cross-linking, pre-hybridization and hybridization was done as described elsewhere in [Farrell, J.R.E., RNA Methodologies. A laboratory Guide for isolation and characterization. San Diego, USA: Academic Press (1993)]. The washing conditions used were: 2 x 20 min. in 2xSSPE/0.1% SDS followed by 1 x 20 min. in 0.1% SSPE/0.1% SDS at 65°C. Northern blots were then analyzed either by a Phosphorimager (Molecular Dynamics) or by autoradiography on X-ray films from Kodak.

Isolation of genomic DNA: *B. subtilis* genomic DNA was isolated from 25 ml overnight cultures according to the standard procedure method 2.6 described by [13].

Southern blot analysis: For hybridization experiments *B. subtilis* genomic DNA (3 mg) was digested with the appropriate restriction enzymes and electrophoresed on a 0.75% agarose gel. The transfer to Zeta-Probe blotting membranes (BIO-RAD), was done as described [Southern, E.M., s.a.]. Prehybridization and hybridization was in 7%SDS, 1% BSA (fraction V; Boehringer), 0.5M Na_2HPO_4, pH 7.2 at 65°C. After hybridization the membranes were washed twice for 5 min. in 2x SSC, 1% SDS at room temperature and twice for 15 min. in 0.1% SSC, 0.1% SDS at 65° C. Southern blots were then analyzed either by a Phosphorimager (Molecular Dynamics) or by autoradiography on X-ray films from Kodak.

DNA sequence analysis: The sequence was determined by the dideoxy chain termination technique [Sanger et al., s.a.] using the Sequenase Kit Version 1.0 (United States Biochemical). Sequence analysis were done using the GCG sequence analysis software package (Version 8.0) by Genetics Computer, Inc. [Devereux et al., s.a.].

Gene amplification in *B. subtilis*: To amplify the copy number of the SFCO in *B. subtilis* transformants, a single colony was inoculated in 15 ml VY-medium supplemented with 1.5 % glucose and 0.02 mg chloramphenicol or neomycin/ml, dependend on the antibiotic resistance gene present in the amplifiable structure (see results and discussion). The next day 750 ml of this culture were used to inoculate 13 ml VY-medium containing 1.5% glucose supplemented with (60, 80, 120 and 150 mg/ml) for the cat resistant mutants, or 160 mg/ml and 180 mg/ml for the neomycin resistant mutants). The cultures were grown overnight and the next day 50 ml of different dilutions (1: 20, 1:400, 1: 8000, 1: 160'000) were plated on VY agar plates with the appropriate antibiotic concentration. Large single colonies were then further analyzed to determine the number of copies and the amount of carotenoids produced.

Analysis of carotenoids: *E. coli* or *B. subtilis* transformants (200 - 400 ml) were grown for the times indicated in the text, usually 24 to 72 hours, in LB-medium or VY-medium, respectively, supplemented with antibiotics, in shake flasks at 37° C and 220 rpm.

The carotenoids produced by the microorganisms were extracted with an adequate volume of acetone using a rotation homogenizer (Polytron, Kinematica AG, CH-Luzern). The homogenate was the filtered through the sintered glass of a suction filter into a round bottom flask. The filtrate was evaporated by means of a rotation evaporator at 50° C using a water-jet vacuum. For the zeaxanthin detection the residue was dissolved in n-hexane/acetone (86:14) before analysis with a normalphase HPLC as described in [Weber, S., s.a.]. For the detection of β-carotene and lycopene the evaporated extract was dissolved in n-hexane/acetone (99:1) and analysed by HPLC as described in Hengartner et al., s.a.].

Example 4

Carotenoid production in *E. coli*

The biochemical assignment of the gene products of the different open reading frames (ORF's) of the carotenoid biosynthesis cluster of *Flavobacterium sp.* were revealed by analyzing the carotenoid accumulation in *E. coli* host strains, transformed with plasmids carrying deletions of the *Flavobacterium sp.* gene cluster, and thus lacking some of the crt gene products. Similar functional assays in *E. coli* have been described by other authors [Misawa et al., s.a.; Perry et al., J. Bacteriol., 168, 607-612 (1986); Hundle, et al., Molecular and General Genetics 254 (4), 406-416 (1994)]. Three different plasmid pLyco, pBIIKS(+)-clone59-2 and pZea4 were constructed from the three genomic isolates pBIIKS(+)-clone2, pBIIKS(+)-clone59 and pBIIKS(+)-clone6a (see figure 16).

Plasmid pBIIKS(+)-clone59-2 was obtained by subcloning the HindIII/BamHI fragment of pBIIKS(+)-clone 2 into the HindIII/BamHI sites of pBIIKS(+)-clone59. The resulting plasmid pBIIKS(+)-clone59-2 carries the complete ORF's of the crtE, crtB, crtI and crtY gene and should lead to the production of β-carotene. pLyco was obtained by deleting the KpnI/KpnI fragment, coding for approx. one half (N-terminus) of the crtY gene, from the plasmid pBIIKS(+)-clone59-2. *E. coli* cells transformed with pLyco, and therefore having a truncated non-functional crtY gene, should produce lycopene, the precursor of β-carotene. pZea4 was constructed by ligation of the AscI-SpeI fragment of pBIIKS(+)-clone59-2, containing the crtE, crtB, crtI and most of the crtY gene with the AscI/XbaI fragment of clone 6a, containing the crtZ gene and sequences to complete the truncated crtY gene mentioned above. pZea4 has therefore all five ORF's of the zeaxanthin biosynthesis pathway. *E. coli* cells transformed with this latter plasmid should therefore produce zeaxanthin. For the detection of the carotenoid produced, transformants were grown for 43 hours in shake flasks and then subjected to carotenoid analysis as described in the methods section. Figure 16 summarizes the construction of the plasmids described above.

As expected the pLyco carrying *E. coli* cells produced lycopene, those carrying pBIIKS(+)-clone59-2 produced β-carotene (all-E,9-Z,13-Z) and the cells having the pZea4 construct produced zeaxanthin. This confirms that we have cloned all the necessary genes of *Flavobacterium sp.* R1534 for the synthesis of zeaxanthin or their precursors (phytoene, lycopene and β-carotene). The production levels obtained are shown in table 1.

plasmid	host	zeaxanthin	β-carotene	lycopene
pLyco	*E. coli* JM109	ND	ND	0.05%
pBIIKS(+)-clone59-2	"	ND	0.03%	ND
pZea4	"	0.033%	0.0009%	ND

Table 1: Carotenoid content of *E. coli* transformants, carrying the plasmids pLyco, pBIIKS(+)-clone59-2 and pZea4, after 43 hours of culture in shake flasks. The values indicated show the carotenoid content in % of the total dry cell mass (200 ml). ND = not detectable.

Examples 5

Carotenoid production in *B. subtilis*

In a first approach to produce carotenoids in *B. subtilis*, we cloned the carotenoid biosynthesis genes of *Flavobacterium* into the Gram (+)/(-) shuttle vectors p602/22, a derivative of p602/20 [LeGrice, S.F.J., s.a.]. The assembling of the final construct p602-CARVEG-E, begins with a triple ligation of fragments PvuII-AvrII of pZea4(del654-

3028) and the AvrII-EcoRI fragment from plasmid pBIIKS(+)-clone6a, into the EcoRI and ScaI sites of the vector p602/22. The plasmid pZea4(del654-3028) had been obtained by digesting pZea4 with SacI and EspI. The protruding and recessed ends were made blunt with Klenow enzyme and religated. Construct pZea4(del654-3028) lacks most of the sequence upstream of crtE gene, which are not needed for the carotenoid biosynthesis. The plasmid p602-CAR has approx. 6.7 kb of genomic *Flavobacterium R1534* DNA containing besides all five carotenoid genes (approx. 4.9 kb), additional genomic DNA of 1.2 kb, located upstream of the crtZ translation start site and further 200 bp, located upstream of crtE transcription start. The crtZ, crtY, crtI and crtB genes were cloned downstream of the $P_{N25/0}$ promoter, a regulatable *E. coli* bacteriophage T5 promoter derivative, fused to a lac operator element, which is functional in *B. subtilis* [LeGrice, S.F.J., s.a.]. It is obvious that in the p602CAR construct, the distance of over 1200 bp between the $P_{N25/0}$ promoter and the transcription start site of crtZ is not optimal and will be improved at a later stage. An outline of the p602CAR construction is shown in figure 17. To ensure transcription of the crtE gene in *B. subtilis*, the vegI promoter [Moran et al., Mol. Gen. Genet. <u>186</u>, 339-346 (1982); LeGrice et al., Mol. Gen. Genet. <u>204</u>, 229-236 (1986)] was introduced upstream of this gene, resulting in the plasmid construct p602-CARVEG-E. The vegI promoter, which originates from siteI of the veg promoter complex described by [LeGrice et al., s.a.] has been shown to be functional in *E. coli* [Moran et al., s.a.]. To obtain this new construct, the plasmid p602CAR was digested with SalI and HindIII, and the fragment containing the complete crtE gene and most of the crtB coding sequence, was subcloned into the XhoI and HindIII sites of plasmid p205. The resulting plasmid p205CAR contains the crtE gene just downstream of the PvegI promoter. To reconstitute the carotenoid gene cluster of *Flavobacterium sp.* the following three pieces were isolated: PmeI/HindIII fragment of p205CAR, the HincII/XbaI fragment and the EcoRI/HindIII fragment of p602CAR and ligated into the EcoRI and XbaI sites of pBluescriptIIKS(+), resulting in the construct pBIIKS(+)-CARVEG-E. Isolation of the EcoRI-XbaI fragment of this latter plasmid and ligation into the EcoRI and XbaI sites of p602/22 gives a plasmid similar to p602CAR but having the crtE gene driven by the PvegI promoter. All the construction steps to get the plasmid p602CARVEG-E are outlined in figure 18. *E. coli* TG1 cells transformed with this plasmid synthesized zeaxanthin. In contrast *B. subtilis* strain 1012 transformed with the same constructs did not produce any carotenoids. Analysis of several zeaxanthin negative *B. subtilis* transformants always revealed, that the transformed plasmids had undergone severe deletions. This instability could be due to the large size of the constructs.

In order to obtain a stable construct in *B. subtilis*, the carotenoid genes were cloned into the Gram (+)/(-) shuttle vector pHP13 constructed by [Haima et al., s.a.]. The stability problems were thought to be omitted by 1) reducing the size of the cloned insert which carries the carotenoid genes and 2) reversing the orientation of the crtE gene and thus only requiring one promoter for the expression of all five genes, instead of two, like in the previous constructs. Furthermore, the ability of cells transformed by such a plasmid carrying the synthetic *Flavobacterium* carotenoid operon (SFCO), to produce carotenoids, would answer the question if a modular approach is feasible. Figure 19 summarizes all the construction steps and intermediate plasmids made to get the final construct pHP13-2PNZYIB-EINV. Briefly: To facilitate the following constructions, a vector pHP13-2 was made, by introducing a synthetic linker obtained with primer CS1 and CS2, between the HindIII and EcoRI sites of the shuttle vector pHP13. The intermediate construct pHP13-2CARVEG-E was constructed by subcloning the AflII-XbaI fragment of p602CARVEG-E into the AflII and XbaI sites of pHP13-2. The next step consisted in the inversion of crtE gene, by removing XbaI and AvrII fragment containing the original crtE gene and replacing it with the XbaI-AvrII fragment of plasmid pBIIKS(+)-PCRRBScrtE. The resulting plasmid was named pHP13-2CARZYIB-EINV and represented the first construction with a functional SFCO. The intermediate construct pBIIKS(+)-PCRRBScrtE mentioned above, was obtained by digesting the PCR product generated with primers #100 and #101 with SpeI and SmaI and ligating into the SpeI and SmaI sites of pBluescriptIIKS(+). In order to get the crtZ transcription start close to the promoter $P_{N25/0}$ a triple ligation was done with the BamHI-SalI fragment of pHP13-2CARZYIB-EINV (contains four of the five carotenoid genes), the BamHI-EcoRI fragment of the same plasmid containing the $P_{N25/0}$ promoter and the EcoRI-SalI fragment of pBIIKS(+)-PCRRBScrtZ, having most of the crtZ gene preceded by a synthetic RBS. The aforementioned plasmid pBIISK(+)-PCRRBScrtZ was obtained by digesting the PCR product amplified with primers #104 and #105 with EcoRI and SalI and ligating into the EcoRI and SalI sites of pBluescriptIISK(+). In the resulting vector pHP13-2PN25ZYIB-EINV, the SFCO is driven by the bacteriophage T5 promoter $P_{N25/0}$, which should be constitutively expressed, due to the absence of a functional lac repressor in the construct [Peschke and Beuk, J. Mol. Biol. <u>186</u>, 547-555 (1985)]. *E. coli* TG1 cells transformed with this construct produced zeaxanthin. Nevertheless, when this plasmid was transformed into *B. subtilis*, no carotenoid production could be detected. Analysis of the plasmids of these transformants showed severe deletions, pointing towards instability problems, similar to the observations made with the aforementioned plasmids.

<u>Examples 6</u>

<u>Chromosome Integration Constructs</u>

Due to the instability observed with the previous constructs we decided to integrate the carotenoid biosynthesis

genes of *Flavobacterium sp.* into the genome of *B. subtilis* using the integration/expression vector pXI12. This vector allows the constitutive expression of whole operons after integration into the levan-sucrase gene (sacB) of the *B. subtilis* genome. The constitutive expression is driven by the vegI promoter and results in medium level expression. The plasmid pXI12-ZYIB-EINV4 containing the synthetic *Flavobacterium* carotenoid operon (SFCO) was constructed as follows: the NdeI-HincII fragment of pBIISK(+)-PCRRBScrtZ was cloned into the NdeI and SmaI sites of pXI12 and the resulting plasmid was named pXI12-PCRcrtZ. In the next step, the BstEII-PmeI fragment of pHP13-2PN25ZYIB-EINV was ligated to the BstEII-PmeI fragment of pXI12-PCRcrtZ (see figure 20). *B. subtilis* transformed with the resulting construct pXI12-ZYIB-EINV4 can integrate the CAR genes either via a Campbell type reaction or via a reciprocal recombination. One transformant, BS1012::ZYIB-EINV4, having a reciprocal recombination of the carotenoid biosynthesis genes into the levan-sucrase gene was further analyzed (figure 21). Although this strain did not synthesize carotenoids, RNA analysis by Northern blots showed the presence of specific polycistronic mRNA's of 5.4 kb and 4.2 kb when hybridized to probe A (see figure 21, panel B). Whereas the larger mRNA has the expected message size, the origin of the shorter mRNA was unclear. Hybridization of the same Northern blot to probe B only detected the large mRNA fragment, pointing towards a premature termination of the transcription at the end of the crtB gene. The presence of a termination signal at this location would make sense, since in the original operon organisation in the *Flavobacterium sp. R1534* genome, the crtE and the crtB genes are facing each other. With this constellation a transcription termination signal at the 5' end of crtB would make sense, in order to avoid the synthesis of anti-sense RNA which could interfere with the mRNA transcript of the crtE gene. Since this region has been changed considerably with respect to the wild type situation, the sequences constituting this terminator may also have been altered resulting in a "leaky" terminator. Western blot analysis using antisera against the different crt enzymes of the carotenoid pathway, pointed towards the possibility that the ribosomal binding sites might be responsible for the lack of carotenoid synthesis. Out of the five genes introduced only the product of crtZ, the β-carotene hydroxylase was detectable. This is the only gene preceded by a RBS site, originating from the pXI12 vector, known to be functional in *B. subtilis*. Base pairing interactions between a mRNA's Shine-Dalgarno sequence [Shine and Delagarno, s. a.] and the 16S rRNA, which permits the ribosome to select the proper initiation site, have been proposed by [McLaughlin et al., J. Biol. Chem. 256, 11283-11291 (1981)] to be much more stable in Gram-positive organisms (*B. subtilis*) than in Gram-negative organisms (*E. coli*). In order to obtain highly stable complexes we exchanged the RBS sites of the Gram-negative *Flavobacterium sp.*, preceding each of the genes crtY, crtI, crtB and crtE, with synthetic RBS's which were designed complementary to the 3' end of the *B. subtilis* 16S rRNA (see table 2). This exchange should allow an effective translation initiation of the different carotenoid genes in *B. subtilis*. The strategy chosen to construct this pXI12-ZYIB-EINV4MUTRBS2C, containing all four altered sites is summarized in figure 20. In order to facilitate the further cloning steps in pBluescriptIIKS(+), additional restriction sites were introduced using the linker obtained with primer MUT7 and MUT8, cloned between the SalI and HindIII sites of said vector. The new resulting construct pBIIKS(+)-LINKER78 had the following restriction sites introduced: AvrII, PmlI, MulI, MunI, BamHI and SphI. The general approach chosen to create the synthetic RBS's upstream of the different carotenoid genes, was done using a combination of PCR based mutagenesis, where the genes were reconstructed using defined primers carrying the modified RBS sites, or using synthetic linkers having such sequences. Reconstitution of the RBS preceding the crtI and crtB genes was done by amplifying the crtI gene with the primers MUT2 and MUT6, which include the appropriate altered RBS sites. The PCR-I fragment obtained was digested with MunI and BamHI and ligated into the MunI and BamHI sites of pBIIKS(+)-LINKER78. The resulting intermediate construct was named pBIIKS(+)-LINKER78PCRI. Reconstitution of the RBS preceding the crtB gene was done using a small PCR fragment obtained with primer MUT3, carrying the altered RBS site upstream of crtB, and primer CAR17. The amplified PCR-F fragment was digested with BamHI and HindIII and sub cloned into the BamHI and HindIII sites of pBIIKS(+)-LINKER78, resulting in the construct pBIIKS(+)-LINKER78PCRF. The PCR-I fragment was cut out of pBIIKS(+)-LINKER78PCRI with BamHI and SapI and ligated into the BamHI and SapI sites of pBIIKS(+)-LINKER78PCRF. The resulting plasmid pBIIKS(+)-LINKER78PCRFI has the PCR-I fragment fused to the PCR-F fragment. This construct was cut with SalI and PmlI and a synthetic linker obtained by annealing of primer MUT9 and MUT10 was introduced. This latter step was done to facilitate the upcoming replacement of the original *Flavobacterium* RBS in the above mentioned construct. The resulting plasmid was named pBIIKS(+)-LINKER78PCRFIA. Assembling of the synthetic RBS's preceding the crtY and crtI genes was done by PCR, using primers MUT1 and MUT5. The amplified fragment PCR-G was made blunt end before cloning into the SmaI site of pUC18, resulting in construct pUC18-PCR-G. The next step was the cloning of the PCR-G fragment between the PCR-A and PCR-I fragments. For this purpose the PCR-G was isolated from pUC18-PCR-G by digesting with MunI and PmlI and ligated into the MunI and PmlI sites of pBIIKS(+)-LINKER78PCRFIA. This construct contains all four fragments, PCR-F, PCR-I, PCR-G and PCR-A, assembled adjacent to each other and containing three of the four artificial RBS sites (crtY, crtI and crtB). The exchange of the *Flavobacterium* RBS's preceding the genes crtY, crtI and crtB by synthetic ones, was done by replacing the HindIII-SalI fragment of plasmid pXI12-ZYIB-EINV4 with the HindIII-SalI fragment of plasmid pBIIKS(+)-LINKER78PCRFIGA. The resulting plasmid pXI12-ZYIB-EINV4 MUTRBSC was subsequently transformed into *E. coli* TG1 cells and *B. subtilis 1012*. The production of zeaxanthin by these cells confirmed that the PCR amplified genes where functional. The *B.*

subtilis strain obtained was named BS1012::SFCO1. The last *Flavobacterium* RBS to be exchanged was the one preceding the crtE gene. This was done using a linker obtained using primer MUT11 and MUT12. The wild type RBS was removed from pXI12-ZYIB-EINV4MUTRBS with NdeI and SpeI and the above mentioned linker was inserted. In the construct pXI12-ZYIB-EINV4MUTRBS2C all *Flavobacterium* RBS's have been replaced by synthetic RBS's of the consensus sequence AAAGGAGG- 7-8 N -ATG (see table 2). *E. coli* TG1 cells transformed with this construct showed that also this last RBS replacement had not interferred

Table 2

mRNA	nucleotide sequence
crtZ	**AAAGGAGG**GUUUCAUAUGAGC
crtY	**AAAGGAGG**ACACGUGAUGAGC
crtI	**AAAGGAGG**CAAUUGAGAUGAGU
crtB	**AAAGGAGG**AUCCAAUCAUGACC
crtE	**AAAGGAGG**GUUUCUUAUGACG

B. subtilis	16S rRNA	3'-UC**UUUCCUCC**ACUAG
E. coli	16S rRNA	3'- AUUCCUCCACUAG

Table 2: Nucleotide sequences of the synthetic ribosome binding sites in the constructs pXI12-ZYIB-EINV4MUTRBS2C, pXI12-ZYIB-EINV4MUTRBS2CCAT and pXI12-ZYIB-EINV4 MUTRBS2CNEO. Nucleotides of the Shine-Dalgarno sequence preceding the individual carotenoid genes which are complementary to the 3' ends of the 16S rRNA of *B. subtilis* are shown in bold. The 3' ends of the 16S rRNA of *E. coli* is also shown as comparison. The underlined AUG is the translation start site of the mentioned gene.

with the ability to produce zeaxanthin. All the regions containing the newly introduced synthetic RBS's were confirmed by sequencing. *B. subtilis* cells were transformed with plasmid pXI12-ZYIB-EINV4MUTRBS2 and one transformant having integrated the SFCO by reciprocal recombination, into the levan-sucrase gene of the chromosome, was selected. This strain was named BS1012::SFCO2. Analysis of the carotenoid production of this strain show that the amounts zeaxanthin produced is approx. 40% of the zeaxanthin produced by *E. coli* cells transformed with the plasmid used to get the *B. subtilis* transformant. Similar was the observation when comparing the BS1012::SFCO1 strain with its *E. coli* counter part (30%). Although the *E. coli* cells have 18 times more carotenoid genes, the carotenoid production

is only a factor of 2-3 times higher. More drastic was the difference observed in the carotenoid contents, between *E. coli* cells carrying the pZea4 construct in about 200 copies and the *E. coli* carrying the plasmid pXI12-ZYIB-EINV4MUTRBS2C in 18 copies. The first transformant produced 48x more zeaxanthin than the latter one. This difference seen can not only be attributed to the roughly 11 times more carotenoid biosynthesis genes present in these transformants. Contributing to this difference is probably also the suboptimal performance of the newly constructed SFCO, in which the overlapping genes of the wild type *Flavobacterium* operon were separated to introduce the synthetic RBS's. This could have resulted in a lower translation efficiency of the rebuild synthetic operon (e.g. due to elimination of putative translational coupling effects, present in the wild type operon).

In order to increase the carotenoid production, two new constructs were made, pXI12-ZYIB-EINV4MUTRBS2CNEO and pXI12-ZYIB-EINV4 MUTRBS2CCAT, which after the integration of the SFCO into the levan-sucrase site of the chromosome, generate strains with an amplifiable structure as described by [Janniere et al., Gene 40, 47-55 (1985)]. Plasmid pXI12-ZYIB-EINV4MUTRBS2CNEO has been deposited on May 25, 1995 at the DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany) under accession No. DSM 10013. Such amplifiable structures, when linked to a resistance marker (e.g chloramphenicol, neomycin, tetracycline), can be amplified to 20-50 copies per chromosome. The amplifiable structure consist of the SFCO, the resistance gene and the pXI12 sequence, flanked by direct repeats of the sac-B 3' gene (see figure 22). New strains having elevated numbers of the SFCO could now be obtained by selecting for transformants with increased level of resistance to the antibiotic. To construct plasmid pXI12-ZYIB-EINV4MUTRBS2CNEO, the neomycin resistance gene was isolated from plasmid pBEST501 with PstI and SmaI and subcloned into the PstI and EcoO1091 sites of the pUC18 vector. The resulting construct was named pUC18-Neo. To get the final construct, the PmeI - AatII fragment of plasmid pXI12-ZYIB-EINV4MUTRBS2C was replaced with the SmaI-AatII fragment of pUC18-Neo, containing the neomycin resistance gene. Plasmid pXI12-ZYIB-EINV4MUTRBS2CCAT was obtained as follows: the chloramphenicol resistance gene of pC194 was isolated by PCR using the primer pair cat3 and cat4. The fragment was digested with EcoRI and AatII and subcloned into the EcoRI and AatII sites of pUC18. The resulting plasmid was named pUC18-CAT. The final vector was obtained by replacing the PmeI-AatII fragment of pXI12-ZYIB-EINV4MUTRBS2C with the EcoRI-AatII fragment of pUC18-CAT, carrying the chloramphenicol resistance gene. Figure 23 summarizes the different steps to obtain aforementioned constructs. Both plasmids were transformed into *B. subtilis* strain 1012, and transformants resulting from a Campbell-type integration were selected. Two strains BS1012::SFCONEO1 and BS1012::SFCOCAT1 were chosen for further amplification. Individual colonies of both strains were independently amplified by growing them in different concentrations of antibiotics as described in the methods section. For the cat gene carrying strain, the chloramphenicol concentrations were 60, 80, 120 and 150 mg/ml. For the neo gene carrying strain, the neomycin concentrations were 160 and 180 mg/ml. In both strains only strains with minor amplifications of the SFCO's were obtained. In daughter strains generated from strain BS1012::SFCONEO1, the resistance to higher neomycin concentrations correlated with the increase in the number of SFCO's in the chromosome and with higher levels of carotenoids produced by these cells. A different result was obtained with daughter strains obtained from strain BS1012::SFCOCAT1. In these strains an increase up to 150 mg chloramphenicol/ml resulted, as expected, in a higher number of SFCO copies in the chromosome.

Example 7

Construction of CrtW containing plasmids and use for carotenoid production

Polymerase chain reaction based gene synthesis. The nucleotide sequence of the artificial crtW gene, encoding the β-carotene β-4-oxygenase of *Alcaligenes* strain PC-1, was obtained by back translating the amino acid sequence outlined in [Misawa, 1995], using the BackTranslate program of the GCG Wisconsin Sequence Analysis Package, Version 8.0 (Genetics Computer Group, Madison, WI, USA) and a codon frequency reference table of *E. coli* (supplied by the Bach Translate Program). The synthetic gene consisting of 726 nucleotides was constructed basically according to the method described by [Ye, 1992]. The sequence of the 12 oligonucleotides (crtW1 - crtW12) required for the synthesis are shown in Figure 25. Briefly, the long oligonucleotides were designed to have short overlaps of 15-20 bases, serving as primers for the extension of the oligonucleotides. After four cycles a few copies of the full length gene should be present which is then amplified by the two terminal oligonucleotides crtW15 and crtW26. The sequences for these two short oligonucleotides are for the forward primer crtW15 (5'-TATATCTAGAcatatgTCCGGTCGTAAA CCGG-3') and for the reverse primer crtW26 (5'-TATAgaattccacgtgTCA AGCACGACCACCGGTTTTAC G-3'), where the sequences matching the DNA templates are underlined. Small cap letters show the introduced restriction sites (*Nde*I for the forward primer and *Eco*RI and *Pml*I for the reverse primer) for the latter cloning into the pALTER-Ex2 expression vector.

Polymerase chain reaction. All twelve long oligonucleotides (crtW1-crtW12; 7 nM each) and both terminal primers (crtW15 and crtW26; 0.1 mM each) were mixed and added to a PCR reaction mix containing Expand™ High Fidelity polymerase (Boehringer, Mannheim) (3.5 units) and dNTP's (100 mM each). The PCR reaction was run for 30 cycles

with the following profile: 94 °C for 1 min, 50 °C for 2 min and 72 °C for 3 min. The PCR reaction was separated on a 1% agarose gel, and the band of approx. 700 bp was excised and purified using the glass beads method (Geneclean Kit, Bio101, Vista, CA, USA). The fragment was subsequentely cloned into the *Sma*I site of plasmid pUC18, using the Sure-Clone Kit (Pharmacia, Uppsala, Sweden). The sequence of the resulting crtW synthetic gene was verified by sequencing with the Sequenase Kit Version 1.0 (United States Biochemical, Cleveland, OH, USA). The crtW gene constructed by this method was found to contain minor errors, which were subsequently corrected by site-directed mutagenesis.

Construction of plasmids. Plasmid pBIIKS(+)-CARVEG-E (see also Example 5) (Figure 26) contains the carotenoid biosynthesis genes (crtE, crtB, crtY, crtI and crtZ) of the Gram (-) bacterium *Flavobacterium* sp. strain R1534 WT (ATCC 21588) [Pasamontes, 1995 #732] cloned into a modified pBluescript II KS(+) vector (Stratagene, La Jolla, USA) carrying site I of the *B. subtilis* veg promoter [LeGrice, 1986 #806]. This constitutive promoter has been shown to be functional in *E. coli*. Transformants of *E. coli* strain TG1 carrying plasmid pBIIKS(+)-CARVEG-E synthesise zeaxanthin. Plasmid pALTER-Ex2-crtW was constructed by cloning the *Nde*I - *Eco*RI restricted fragment of the synthetic crtW gene into the corresponding sites of plasmid pALTER-Ex2 (Promega, Madison, WI). Plasmid pALTER-Ex2 is a low copy plasmid with the p15a origin of replication, which allows it to be maintained with ColE1 vectors in the same host. Plasmid pBIIKS-crtEBIYZW (Figure 26) was obtained by cloning the *Hind*III-*Pml*I fragment of pALTER-Ex2-crtW into the *Hind*III and the blunt end made *Mlu*I site obtained by a fill in reaction with Klenow enzyme, as described elsewhere in [Sambrook, 1989 #505]. Inactivation of the crtZ gene was done by deleting a 285 bp *Nsi*I-*Nsi*I fragment, followed by a fill in reaction and religation, resulting in plasmid pBIIKS-crtEBIY[DZ]W. Plasmid pBIIKS-crtEBIY[DZW] carrying the non-functional genes crtW and crtZ, was constructed by digesting the plasmid pBIIKS-crtEBIY[DZ]W with *Nde*I and *Hpa*I, and subsequent self religation of the plasmid after filling in the sites with Klenow enzyme. *E. coli* transformed with this plasmid had a yellow-orange colour due to the accumulation of β-carotene. Plasmid pBIIKS-crtEBIYZ[DW] has a truncated crtW gene obtained by deleting the *Nde*I - *Hpa*I fragment in plasmid pBIIKS-crtEBIYZW as outlined above. Plasmids pALTER-Ex2-crtEBIY[DZW] and pALTER-Ex2-crtEBIYZ[DW], were obtained by isolating the *Bam*HI-*Xba*I fragment from pBIIKS-crtEBIY[DZW] and pBIIKS-crtEBIYZ[DW], respectively and cloning them into the *Bam*HI and *Xba*I sites of pALTER-Ex2. The plasmid pBIIKS-crtW was constructed by digesting pBIIKS-crtEBIYZW with *Nsi*I and *Sac*I, and self-religating the plasmid after recessing the DNA overhangs with Klenow enzyme. Figure 27 compiles the relevant inserts of all the plasmids used in this paper.

Carotenoid analysis. *E. coli* TG-1 transformants carrying the different plasmid constructs were grown for 20 hours in Luria-Broth medium supplemented with antibiotics (ampicillin 100 mg/ml, tetracyclin 12.5 mg/ml) in shake flasks at 37°C and 220 rpm. Carotenoids were extracted from the cells with acetone. The acetone was removed in vacuo and the residue was re dissolved in toluene. The coloured solutions were subjected to high-performance liquid chromatography (HPLC) analysis which was performed on a Hewlett-Packard series 1050 instrument. The carotenoids were separated on a silica column Nucleosil Si - 100, 200 x 4 mm, 3m. The solvent system included two solvents: hexane (A) and hexane/THF , 1:1 (B). A linear gradient was applied running from 13 to 50 % (B) within 15 minutes. The flow rate was 1.5 ml/min. Peaks were detected at 450 nm by a photo diode array detector. The individual carotenoid pigments were identified by their absorption spectra and typical retention times as compared to reference samples of chemically pure carotenoids, prepared by chemical synthesis and characterised by NMR, MS and UV-Spectra. HPLC analysis of the pigments isolated from *E. coli* cells transformed with plasmid pBIIKS-crtEBIYZW, carrying besides the carotenoid biosynthesis genes of *Flavobacterium* sp. strain R1534, also the crtW gene encoding the β-carotene ketolase of *Alcaligenes* PC-1 [Misawa, 1995 #670] gave the following major peaks identified as: b-cryptoxanthin, astaxanthin, adonixanthin and zeaxanthin, based on the retention times and on the comparison of the absorbance spectra to given reference samples of chemically pure carotenoids. The relative amount (area percent) of the accumulated pigment in the E. coli transformant carrying pBIIKS-crtEBIYZW is shown in Table 3 ["CRX": cryptoxanthin; "ASX": astaxanthin; "ADX": adonixanthin; "ZXN": zeaxanthin; "ECM": echinenone; "MECH": 3-hydroxyechinenone, "CXN": cantaxanthin]. The Σ of the peak areas of all identified carotenoids was defined as 100%. Numbers shown in Table 3 represent the average value of four independent cultures for each transformant. In contrast to the aforementioned results, *E. coli* transformants carrying the same genes but on two plasmids namely, pBIIKS-crtEBIYZ[DW] and pALTER-Ex2-crtW, showed a drastical drop in adonixanthin and a complete lack of astaxanthin pigments (Table 3), whereas the relative amount of zeaxanthin (%) had increased. Echinenone, hydroxyechinenone and canthaxanthin levels remained unchanged compared to the transformants carrying all the crt genes on the same plasmid (pBIIKS-crtEBIYZDW). Plasmid pBIIKS-crtEBIYZ[DW] is a high copy plasmid carrying the functional genes of crtE, crtB, crtY, crtI, crtZ of *Flavobacterium* sp. strain R1534 and a truncated, non-functional version of the crtW gene, whereas the functional copy of the crtW gene is located on the low copy plasmid pALTER-Ex2-crtW. To analyze the effect of overexpression of the crtW gene with respect to the crtZ gene, *E. coli* cells were co-transformed with plasmid pBIIKS-crtW carrying the crtW gene on the high copy plasmid pBIIKS-crtW and the low copy construct pALTER-Ex2-crtEBIYZ[DW], encoding the *Flavobacterium* crt genes. Pigment analysis of these transformants by HPLC monitored the presence of β-carotene, cryptoxanthin, astaxanthin, adonixanthin, zeaxanthin, 3-hydroxyechine-none and minute traces of echinenone and canthaxanthin (Table 3).

Transformants harbouring the crtW gene on the low copy plasmid pALTER-Ex2-crtW and the genes crtE, crtB, crtY and crtI on the high copy plasmid pBIIKS-crtEBIY[DZW] expressed only minor amounts of canthaxanthin (*6* %) but high levels of echinenone (*94*%), whereas cells carrying the crtW gene on the high copy plasmid pBIIKS-crtW and the other crt genes on the low copy construct pALTER-Ex2-crtEBIY[DZW], had *78.6* % and *21.4* % of echinenone and canthaxanthin, respectively (Table 3).

Table 3

plasmids	CRX	ASX	ADX	ZXN	ECH	HECH	CXN
pBIIKS-crtEBIYZW	1.1	2.0	44.2	52.4	< 1	< 1	< 1
pBIIKS-crtEBIYZ[ΔW] + pALTER-Ex2-crtW	2.2	-	25.4	72.4	< 1	< 1	< 1
pBIIKS-crtEBIY[ΔZ]W	-	-	-	-	66.5	-	33.5
pBIIKS-crtEBIY[ΔZW] + pBIIKS-crtW	-	-	-	-	94	-	6

Example 8

Selective carotenoid production by using the crtW and crtZ genes of the Gram negative bacterium E-396.

In this section we describe *E. coli* transformants which accumulate only one (canthaxanthin) or two main carotenoids (astaxanthin, adonixanthin) and minor amounts of adonirubin, rather than the complex variety of carotenoids seen in most carotenoid producing bacteria [Yokoyama et al., Biosci. Biotechnol. Biochem. 58:1842-1844 (1994)] and some of the *E. coli* transformants shown in Table 3. The ability to construct strains producing only one carotenoid is a major step towards a successful biotechnological carotenoid production process. This increase in the accumulation of individual carotenoids accompanied by a decrease of the intermediates, was obtained by replacing the crtZ of *Flavobacterium* R1534 and/or the synthetic crtW gene (see example 5) by their homologous genes originating from the astaxanthin producing Gram negative bacterium E-396 (FERM BP-4283) [Tsubokura et al., EP-application 0 635 576 A1]. Both genes, $crtW_{E396}$ and $crtZ_{E396}$, were isolated and used to construct new plasmids as outlined below.

Isolation of a putative fragment of the crtW gene of strain E-396 by the polymerase chain reaction. Based on protein sequence comparison of the crtW enzymes of *Agrobacterium aurantiacum*, *Alcaligenes PC-1* (WO95/18220) [Misawa et al., J.Bacteriol. 177: 6575-6584 (1995)] and *Haematococcus pluvialis* [Kajiwara et al., Plant Mol. Biol. 29:343-352 (1995)][Lotan et al., FEBS letters, 364:125-128 (1995)], two regions named I and II, having high amino acid conservation and located approx. 140 amino acids appart, were identified and chosen to design the degenerate PCR primers shown below. The N-terminal peptide HDAMHG (region I) was used to design the two 17-mer degenerate primer sequences crtW100 and crtW101:

crtW100: 5'-CA(C/T)GA(C/T)GC(A/C)ATGCA(C/T)GG-3'

crtW101: 5'-CA(C/T)GA(C/T)GC(G/T)ATGCA(C/T)GG-3'

The C-terminal peptide H(W/H)EHH(R/L) corresponding to region II was used design the two 17-mer degenerate primer with the antisense sequences crtW105 and crtW106:

crtW105: 5'-AG(G/A)TG(G/A)TG(T/C)TC(G/A)TG(G/A)TG-3'

crtW106: 5'-AG(G/A)TG(G/A)TG(T/C)TCCCA(G/A)TG-3'

Polymerase chain reaction. PCR was performed using the GeneAmp Kit (Perkin Elmer Cetus) according to the manufacturer's instructions. The different PCR reactions contained combinations of the degenerate primers (crtW100/crtW105 or crtW100/crtW106 or crtW101/crtW105 or crtW101/crtW106) at a final concentration of 50 pM each, together with genomic DNA of the bacterium E-396 (200 ng) and 2.5 units of Taq polymerase. In total 35 cycles of PCR were performed with the following cycle profile: 95 °C for 30 sec, 55 °C for 30 sec, 72 °C for 30 sec. PCR reactions made with the following primer combinations crtW100/crtW105 and crtW101/crtW105 gave PCR amplification products of approx. 500 bp which w¯re in accordance with the exp¯cted fragment size. The 500 bp fragment, JAPclone8, obtained in the PCR reaction using primers crtW101 and crtW105 was excised from an 1.5% agarose gel and purified using the GENECLEAN Kit and subsequently cloned into the *Sma*I site of pUC18 using the Sure-Clone Kit,

according to the manufacturer's instructions. The resulting plasmid was named pUC18-JAPclone8 and the insert was sequenced. Comparison of the determined sequence to the crtW gene of *Agrobacterium aurantiacum* (GenBank accession n° D58420) published by Misawa *et al.* in 1995 (WO95/18220) showed 96% identity at the nucleotide sequence level, indicating that both organisms might be closely related.

Isolation of the crt cluster of the strain E- 396. Genomic DNA of E-396 was digested overnight with different combinations of restrictions enzymes and separated by agarose gel electrophoresis before transferring the resulting fragments by Southern blotting onto a nitrocellulose membrane. The blot was hybridised with a ^{32}P labelled 334 bp fragment obtained by digesting the aforementioned PCR fragment JAPclone8 with *BssHII* and *MluI*. An approx. 9,4kb *EcoRI/BamHI* fragment hybridizing to the probe was identified as the most appropiate for cloning since it is long enough to potentially carry the complete *crt* cluster. The fragment was isolated and cloned into the *EcoRI* and *BamHI* sites of pBluescriptIIKS resulting in plasmid pJAPCL544 (Fig. 29). Based on the sequence of the PCR fragment JAPclone8, two primers were synthesized to obtain more sequence information left and right hand of this fragment. Fig. 30 shows the sequence obtained containing the crtW$_{E396}$ (from nucleotide 40 to 768) and crtZ$_{E396}$ (from nucleotide 765 to 1253) genes of the bacterium E-396. The nucleotide sequence of the crtW$_{E396}$ gene is shown in Fig. 31 and the encoded amino acid sequence in Fig. 32. The nucleotide sequence of the crtZ$_{E396}$ gene is shown in Fig. 33 and the corresponding amino acid sequence in Fig. 34. Comparison to the crtW$_{E396}$ gene of E-396 to the crtW gene of *A. aurantiacum* showed 97 % identity at the nucleotide level and 99 % identity at the amino acid level. For the crtZ gene the values were 98 % and 99 %, respectively.

Construction of plasmids: Both genes, crtW$_{E396}$ and crtZ$_{E396}$, which are adjacent in the genome of E-393, were isolated by PCR using primer crtW107 and crtW108 and the ExpandTM High Fidelity PCR system of Boehringer Mannheim, according to the manufacturer's recommendations. To facilitate the subsequent cloning steps (see section below) the primer crt107 (5'-ATCATATGAGCGCACATGCCCTGCCCAAGGC-3') contains an artificial *NdeI* site (underlined sequence) spanning the ATG start codon of the crtW$_{E396}$ gene and the reverse primer crtW108 (5'-ATCTCGAGTCACGTGCGCTCCTGCGCCTCGGCC-3') has an *XhoI* site (underlined sequence) just downstream of the TGA stop codon of the crtZ $_{E396}$ gene. The final PCR reaction mix had 10 pM of each primer, 2.5 mg genomic DNA of the bacterium E-396 and 3.5 units of the TaqDNA/Pwo DNA polymerase mix. In total 35 cycles were performed with the following cycle profile: 95 °C, 1 min; 60 °C, 1 min; 72 °C 1min 30 sec. The PCR product of approx. 1250bp was isolated from the 1% agarose gel and purified using GENECLEAN before ligation into the *SmaI* site pUC18 using the Sure-Clone Kit. The resulting construct was named pUC18-E396crtWZPCR (Fig. 35). The functionality of both genes was tested as follows. The crtW$_{E396}$ and crtZ $_{E396}$ gene were isolated from plasmid pUC18-E396crtWZPCR with *NdeI* and *XhoI* and cloned into the *NdeI* and *SalI* site of plasmid pBIIKS-crtEBIYZW resulting in plasmid pBIIKS-crtEBIY[E396WZ] (Fig. 36). *E. coli* TG1 cells transformed with this plasmid produced astaxanthin, adonixanthin and adonirubin but no zeaxanthin (Table 4).

Plasmid pBIIKS-crtEBIY[E396W]DZ has a truncated non-functional crtZ gene. Fig. 37 outlines the construction of this plasmid. The PCR reaction was run as outlined elsewhere in the text using primers crtW113/crtW114 and 200 ng of plasmid pUC18-JAPclone8 as template using 20 cycles with the following protocol: 95 °C, 45 sec/ 62 °C, 20 sec/ 72 °C, 20 sec)

primer crtW113 (5'-ATATACATATGGTGTCCCCCTTGGTGCGGGTGC-3')

primer crtW114 (5'-TATGGATCCGACGCGTTCCCGGACCGCCACAATGC-3')

The resulting 150 bp fragment was digested with *BamHI* and *NdeI* and cloned into the corresponding sites of pBIISK(+)-PCRRBScrtZ resulting in the construct pBIISK(+)-PCRRBScrtZ-2.The final plasmid carrying the genes crtE, crtB, crtI, crtY of *Flavobacterium*, the crtW$_{E396}$ gene of E-396 and a truncated non-functional crtZ gene of *Flavobacterium* was obtained by isolating the *MluI/NruI* fragment (280 bp) of pBIISK(+)-PCRRBScrtZ-2 and cloning it, into the *MluI/PmlI* sites of plasmid pBIIKS-crtEBIY[E396WZ]. *E. coli* cells transformed with this plasmid produced 100% canthaxanthin (Table 4; "CRX": cryptoxanthin; "ASX": astaxanthin; "ADX": adonixanthin: "ZXN": zeaxanthin; "ECH": echinenone; "HECH": 3-hydroxyechinenone; "CXN": canthaxanthin; "BCA": β-carotene; "ADR": adonirubin; Numbers indicate the % of the individual carotenoid of the total carotenoids produced in the cell.).

Table 4

plasmid	CRX	ASX	ADX	ZXN	ECH	HECH	CXN	BCA	ADR
pBIIKScrtEBIYZW	1.1	2.0	44.2	52.4	<1	<1	<1		
pBIIKS-crtEBIY[E396WZ]		74.4	19.8						5.8

Table 4 (continued)

plasmid	CRX	ASX	ADX	ZXN	ECH	HECH	CXN	BCA	ADR
pBIIKS-crtEBIY[E396W]ΔZ							100		

The results of *E. coli* transformants carrying pBIIKScrtEBIYZW (see example 7) are also shown in Table 4 to indicate the dramatic effect of the new genes crtW$_{E396}$ and crtZ$_{E396}$ on the carotenoids produced in these new transformants.

Example 9

Cloning of the remaining *crt* genes of the Gram negative bacterium E-396.

TG1 *E. coli* transformants carrying the pJAPCL544 plasmid did not produce detectable quantities of carotenoids (results not shown). Sequence analysis and comparison of the 3' (*BamHI* site) of the insert of plasmid pJAPCL544, to the *crt* cluster of *Flavobacterium* R1534 showed that only part of the C-terminus of the crtE gene was present. This result explained the lack of carotenoid production in the aforementioned transformants. To isolate the missing N-terminal part of the gene, genomic DNA of E-396 was digested by 6 restrictions enzymes in different combinations : *Eco*RI, *Bam*HI, *Pst*I, *Sac*I, *Sph*I and *Xba*I and transferred by the Southern blot technique to nitrocellulose. Hybridization of this membrane with the ^{32}P radio-labelled probe (a 463 bp *PstI-BamHI* fragment originating from the 3' end of the insert of pJAPCL544 (Fig. 29) highlighted a ~1300 bp-long *PstI-PstI* fragment. This fragment was isolated and cloned into the *PstI* site of pBSIIKS(+) resulting in plasmid pBSIIKS-#1296. The sequence of the insert is shown in Fig. 38 (small cap letters refer to new sequence obtained. Capital letters show the sequence also present in the 3' of the insert of plasmid pJAPCL544). The complete crtE gene has therefore a length of 882 bp (see Fig. 39) and encodes a GGPP synthase of 294 amino acids (Fig. 40). The crtE enzyme has 38 % identity with the crtE amino acid sequence of *Erwinia herbicola* and 66 % with *Flavobacterium* R1534 WT.

Construction of plasmids. To have a plasmid carrying the complete *crt* cluster of E-396, the 4.7 kb *MluI/Bam*HI fragment encoding the genes crtW, crtZ, crtY, crtI and crtB was isolated from pJAPCL544 and cloned into the *MluI/Bam*HI sites of pUC18-E396crtWZPCR (see example 8). The new construct was named pE396CARcrtW-B (Fig. 41) and lacked the N-terminus of the crtE gene. The missing C-terminal part of the crtE gene was then introduced by ligation of the aforementioned *PstI* fragment of pBIIKS-#1296 between the *PstI* sites of pE396CARcrtW-B. The resulting plasmid was named pE396CARcrtW-E (Fig. 41). The carotenoid distribution of the *E. coli* transformants carrying aforementioned plasmid were: adonixanthin (65%), astaxanthin (8%) and zeaxanthin (3%). The % indicated reflects the proportion of the total amount of carotenoid produced in the cell.

Example 10

Astaxanthin and adonixanthin production *in Flavobacterium* R1534

Among bacteria *Flavobacterium* may represent the best source for the development of a fermentative production process for 3*R*, 3*R*' zeaxanthin. Derivatives of *Flavobacterium* sp. strain R1534, obtained by classical mutagenesis have attracted in the past two decades wide interest for the development of a large scale fermentative production of zeaxanthin, although with little success. Cloning of the carotenoid biosynthesis genes of this organism, as outlined in example 2, may allow replacement of the classical mutagenesis approach by a more rational one, using molecular tools to amplify the copy number of relevant genes, deregulate their expression and eliminate bottlenecks in the carotenoid biosynthesis pathway. Furthermore, the introduction of additional heterologous genes (e.g. crtW) will result in the production of carotenoids normally not synthesised by this bacterium (astaxanthin, adonirubin, adonixanthin, canthaxanthin, echinenone). The construction of such recombinant *Flavobacterium* R1534 strains producing astaxanthin and adonixanthin will be outlined below.

Gene transfer into *Flavobacterium* sp.

Plasmid transfer by conjugative mobilization. For the conjugational crosses we constructed plasmid pRSF1010-Amp^r, a derivative of the small (8.9 kb) broad host range plasmid RSF1010 (IncQ incompatibility group) [Guerry et al., J. Bacteriol. 117:619-630 (1974)] and used E. coli S17-1 as the mobilizing strain [Priefer et al., J. Bacteriol. 163:324-330 (1985)]. In general any of the IncQ plasmids (e.g. RSF1010, R300B, R1162) may be mobilized into rifampicin resistant Flavobacterium *if the* transfer functions are provided by plasmids of the IncP1 group (e.g. R1, R751).

Rifampicin resistant (Rif^r) *Flavobacterium* R1534 cells were obtained by selection on 100 mg rifampicin/ml. One resistant colony was picked and a stock culture was made. The conjugation protocol was as follows:

Day 1:

- grow 3 ml culture of *Flavobacterium* R1534 Rif^r for 24 hours at 30 °C in Flavobacter medium (F-medium) (see example 1)
- grow 3 ml mobilizing *E. coli* strain carrying the mobilizable plasmid O/N at 37 °C in LB medium. (e.g *E. coli* S17-1 carrying pRSF1010-Amp^r or *E. coli* TG-1 cells carrying R751 and pRSF1010-Amp^r)

Day 2:

- pellet 1 ml of the *Flavobacterium* R1534 Rif^r cells and resuspend in 1ml of fresh F-medium.
- pellet 1 ml of *E. coli* cells (see above) and resuspend in 1 ml of LB medium.
- donor and recipient cells are then mixed in a ratio of 1:1 and 1: 10 in an Eppendorf tube and 30 ml are then applied onto a nitrocellulose filter plated on agar plates containing F-medium and incubated O/N at 30°C.

Day 3:

- the conjugational mixtures were washed off with F-medium and plated on F-medium containing 100 mg rifampicin and 100 mg ampicillin/ml for selection of transconjugants and inhibition of the donor cells.

Day 6-8:

- Arising clones are plated once more on F-medium containing 100 mgRif and 100 mg Amp/ml before analysis.

Plasmid transfer by electroporation. The protocol for the eletroporation is as follows:

1. add 10 ml of O/N culture of *Flavobacterium* sp. R1534 into 500 ml F-medium and incubate at 30°C until OD600=0.8-0.1

2. harvest cells by centrifugation at 4000g at 4°C for 10 min.

3. wash cells in equal volume of ice-cold deionized water (2 times)

4. resuspend bacterial pellet in 1 ml ice-cold deionized water

5. take 50 ml of cells for electroporation with 0.1 mg of plasmid DNA

6. electroporation was done using field strengths between 15 and 25 kV/cm and 1-3 ms.

7. after electroporation cells were immediately diluted in 1 ml of F-medium and incubated for 2 hours at 30°C at 180 rpm before plating on F-medium plates containing the respective selective antibioticum.

Plasmid constructions: Plasmid pRSF101-Amp^r was obtained by cloning the Amp^r gene of pBR322 between the *EcoRI/NotI* sites of RSF1010. The Amp^r gene originates from pBR322 and was isolated by PCR using primers AmpR1 and AmpR2 as shown in Fig. 42.

AmpR1:
5'-TATATCGGCCGACTAGTAAGCTTCAAAAAGGATCTTCACCTAG-3' the underlined sequence contains the introduced restriction sites for *EagI*, *SpeI* and *HindIII* to facilitate subsequent constructions.

AmpR2:
5'-ATATGAATTCAATAATATTGAAAAAGGAAG-3' the underlined sequence corresponds to an introduced *EcoRI* restriction site to facilitate cloning into RSF1010 (see Fig. 42).

The PCR reaction mix had 10 pM of each primer (AmpR1/AmpR2), 0.5 mg plasmid pBR322 and 3.5 units of the TaqDNA/Pwo DNA polymerase mix. In total 35 amplification cycles were made with the profile: 95 °C, 45 sec; 59 °C, 45 sec, 72 °C, 1 min. The PCR product of approx. 950 was extracted once with phenol/chloroform and precipitated with 0.3

M NaAcetate and 2 vol. Ethanol. The pellet was resuspended in H_2O and digested with *EcoRI* and *EagI* O/N. The digestion was separated by electrophoresis and the fragment isolated from the 1% agarose gel and purified using GENECLEAN before ligation into the *EcoRI* and *NotI* sites of RSF1010. The resulting plasmid was named pRSF1010-Amp^r (Fig. 42).

Plasmid RSF1010-Ampr-crt1 was obtained by isolating the *HindIII/NotI* fragment of pBIIKS-crtEBIY[E396WZ] and cloning it between the *HindIII*/EagI sites of RSF1010-Amp^r (Fig. 43). The resulting plasmid RSF1010-Ampr-crt1 carries crtW$_{E396}$, crtZ$_{E396}$, crtY genes and the N-terminus of the crtI gene (non-functional). Plasmid RSF1010-Ampr-crt2 carrying a complete crt cluster composed of the genes crtW$_{E396}$ and crtZ$_{E396}$ of E-396 and the crtY, crtI, crtB and crtE of *Flavobacterium* R1534 was obtained by isolating the large *HindIII/XbaI* fragment of pBIIKS-crtEBIY[E396WZ] and cloning it into the *SpeI/HindIII* sites of RSF1010-Amp^r (Fig. 43).

Flavobacterium R1534 transformants carrying either plasmid RSF1010-Amp^r, Plasmid RSF1010-Amp^r-crt1 or Plasmid RSF1010-Amp^r-crt2 were obtained by conjugation as outlined above using *E. coli* S17-1 as mobilizing strain.

Comparison of the carotenoid production of two Flavobacterium transformants. Overnight cultures of the individual transformants were diluted into 20 ml fresh F-medium to have a final starting OD600 of 0.4. Cells were harvested after growing for 48 hours at 30 °C and carotenoid contents were analysed as outlined in example 7. Table 5 shows the result of the three control cultures *Flavobacterium* [R1534 WT], [R1534 WT RifR] (rifampicin resistant) and [R1534WT Rifr RSF1010-AmpR] (carries the RSF1010-Amp^r plasmid) and the two transformants [R1534 WT RSF1010-AmpR-crt1] and [R1534 WT RSF1010-AmpR-crt2]. Both latter transformants are able to synthesise astaxanthin and adonixanthin but little zeaxanthin. Most interesting is the [R1534 WT RSF1010-AmpR-crt2] *Flavobacterium* transformant which produces approx. 4 times more carotenoids than the R1534 WT. This increase in total carotenoid production is most likely due to the increase of the number of carotenoid biosynthesis clusters present in these cell (e.g. corresponds to the total copy number of plasmids in the cell).

Table 5

Transformant	carotenoids % of total dry weight	total carotenoid content in % of dry weight
R1534 WT	0.039% β-Carotin 0.001% β-Cryptoxanthin 0.018% Zeaxanthin	0.06%
R1534 Rif^r	0.036% β-Carotin 0.002% β-Cryptoxanthin 0.022% Zeaxanthin	0.06%
R1534 Rif^r [RSF1010-Ampr]	0.021% β-Carotin 0.002% β-Cryptoxanthin 0.032% Zeaxanthin	0.065%
R1534 Rif^r [RSF1010-Ampr-crt1]	0.022% Astaxanthin 0.075% Adonixanthin 0.004% Zeaxanthin	0.1%
R1534 Rif^r [RSF1010-Ampr-crt2]	0.132% β-Carotin 0.006% Echinenon 0.004% Hydroxyechinenon 0.003% β-Cryptoxanthin 0.044% Astaxanthin 0.039% Adonixanthin 0.007% Zeaxanthin	0.235%

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: F.HOFFMANN-LA ROCHE AG
(B) STREET: GRENZACHERSTRASSE 124
(C) CITY: BASLE
(D) STATE: BS
(E) COUNTRY: SWITZERLAND
(F) POSTAL CODE (ZIP): CH - 4002
(G) TELEPHONE: 061 - 688 2505
(H) TELEFAX: 061 688 1395
(I) TELEX: 962292/965542 hlr ch

(ii) TITLE OF INVENTION: Improved fermentative carotenoid production

(iii) NUMBER OF SEQUENCES: 17

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: EP 97120324.5

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATGAGCGCAC ATGCCCTGCC CAAGGCAGAT CTGACCGCCA CCAGTTTGAT CGTCTCGGGC     60
GGCATCATCG CCGCGTGGCT GGCCCTGCAT GTGCATGCGC TGTGGTTTCT GGACGCGGCG    120
GCGCATCCCA TCCTGGCGGT CGCGAATTTC CTGGGGCTGA CCTGGCTGTC GGTCGGTCTG    180
TTCATCATCG CGCATGACGC GATGCATGGG TCGGTCGTGC CGGGGCGCCC GCGCGCCAAT    240
GCGGCGATGG GCCAGCTTGT CCTGTGGCTG TATGCCGGAT TTTCCTGGCG CAAGATGATC    300
GTCAAGCACA TGGCCCATCA TCGCCATGCC GGAACCGACG ACGACCCAGA TTTCGACCAT    360
GGCGGCCCGG TCCGCTGGTA CGCCCGCTTC ATCGGCACCT ATTTCGGCTG GCGCGAGGGG    420
CTGCTGCTGC CCGTCATCGT GACGGTCTAT GCGCTGATGT TGGGGGATCG CTGGATGTAC    480
GTGGTCTTCT GGCCGTTGCC GTCGATCCTG GCGTCGATCC AGCTGTTCGT GTTCGGCATC    540
TGGCTGCCGC ACCGCCCCGG CCACGACGCG TTCCCGGACC GCCACAATGC GCGGTCGTCG    600
CGGATCAGCG ACCCCGTGTC GCTGCTGACC TGCTTTCACT TTGGCGGTTA TCATCACGAA    660
CACCACCTGC ACCCGACGGT GCCTTGGTGG CGCCTGCCCA GCACCCGCAC CAAGGGGGAC    720
ACCGCATGA                                                            729
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ser Ala His Ala Leu Pro Lys Ala Asp Leu Thr Ala Thr Ser Leu
1               5                   10                  15

Ile Val Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His
            20                  25                  30

Ala Leu Trp Phe Leu Asp Ala Ala Ala His Pro Ile Leu Ala Val Ala
        35                  40                  45

Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala
    50                  55                  60

His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn
65                  70                  75                  80

Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp
                85                  90                  95

Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr
            100                 105                 110

Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala
        115                 120                 125

Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro
    130                 135                 140

Val Ile Val Thr Val Tyr Ala Leu Met Leu Gly Asp Arg Trp Met Tyr
145                 150                 155                 160

Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe
                165                 170                 175

Val Phe Gly Ile Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro
            180                 185                 190

Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu
        195                 200                 205

Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His
    210                 215                 220

Pro Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp
225                 230                 235                 240

Thr Ala
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
ATGACCAATT TCCTGATCGT CGTCGCCACC GTGCTGGTGA TGGAGCTGAC GGCCTATTCC    60
GTCCACCGCT GGATCATGCA CGGCCCCTTG GGCTGGGGCT GGCACAAGTC CCACCACGAG   120
GAACACGACC ACGCGCTGGA AAAGAACGAC CTGTACGGCC TGGTCTTTGC GGTGATCGCC   180
ACGGTGCTGT TCACGGTGGG CTGGATCTGG GCACCGGTCC TGTGGTGGAT CGCCTTGGGC   240
ATGACCGTCT ACGGGCTGAT CTATTTCGTC CTGCATGACG GGCTGGTGCA TCAGCGCTGG   300
CCGTTCCGCT ATATCCCTCG CAAGGGCTAT GCCAGACGCC TGTATCAGGC CCACCGCCTG   360
CACCACGCGG TCGAGGGGCG CGACCATTGC GTCAGCTTCG GCTTCATCTA TGCGCCGCCG   420
GTCGACAAGC TGAAGCAGGA CCTGAAGACG TCGGGCGTGC TGCGGGCCGA GGCGCAGGAG   480
CGCACG                                                              486
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Thr Asn Phe Leu Ile Val Val Ala Thr Val Leu Val Met Glu Leu
1               5                   10                  15

Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro Leu Gly Trp
            20                  25                  30

Gly Trp His Lys Ser His His Glu Glu His Asp His Ala Leu Glu Lys
        35                  40                  45

Asn Asp Leu Tyr Gly Leu Val Phe Ala Val Ile Ala Thr Val Leu Phe
    50                  55                  60

Thr Val Gly Trp Ile Trp Ala Pro Val Leu Trp Trp Ile Ala Leu Gly
65                  70                  75                  80

Met Thr Val Tyr Gly Leu Ile Tyr Phe Val Leu His Asp Gly Leu Val
                85                  90                  95

His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr Ala Arg
            100                 105                 110

Arg Leu Tyr Gln Ala His Arg Leu His His Ala Val Glu Gly Arg Asp
        115                 120                 125

His Cys Val Ser Phe Gly Phe Ile Tyr Ala Pro Pro Val Asp Lys Leu
    130                 135                 140

Lys Gln Asp Leu Lys Thr Ser Gly Val Leu Arg Ala Glu Ala Gln Glu
145                 150                 155                 160

Arg Thr
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 882 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
ATGAGACGAG ACGTCAACCC GATCCACGCC ACCCTTCTGC AGACCAGACT TGAGGAGATC    60
GCCCAGGGAT TCGGTGCCGT GTCGCAGCCG CTCGGCCCGG CCATGAGCCA TGGCGCGCTG   120
TCGTCGGGCA AGCGTTTCCG CGGCATGCTG ATGCTGCTTG CGGCAGAAGC CTCGGGCGGG   180
GTCTGCGACA CGATCGTCGA CGCCGCCTGC GCGGTCGAGA TGGTGCATGC CGCATCGCTG   240
ATCTTCGACG ACCTGCCCTG CATGGACGAT GCCGGGCTGC GCCGCGGCCA GCCCGCGACC   300
CATGTGGCGC ATGGCGAAAG CCGCGCCGTG CTAGGCGGCA TCGCCCTGAT CACCGAGGCG   360
ATGGCCCTGC TGGCCGGTGC GCGCGGCGCG TCGGGCACGG TGCGGGCGCA GCTGGTGCGG   420
ATCCTGTCGC GGTCCCTGGG GCCGCAGGGC CTGTGCGCCG GCCAGGACCT GGACCTGCAC   480
GCGGCCAAGA ACGGCGCGGG GGTCGAACAG GAACAGGACC TGAAGACCGG CGTGCTGTTC   540
ATCGCCGGGC TGGAGATGCT GGCCGTGATC AAGGAGTTCG ACGCCGAGGA GCAGACTCAG   600
ATGATCGACT TTGGCCGTCA GCTGGGCCGG GTGTTCCAGT CCTATGACGA CCTGCTGGAC   660
GTTGTGGGCG ACCAGGCGGC GCTTGGCAAG GATACCGGTC GCGATGCGGC GGCCCCCGGC   720
CCGCGGCGCG GCCTTCTGGC CGTGTCAGAC CTGCAGAACG TGTCCCGTCA CTATGAGGCC   780
AGCCGCGCCC AGCTGGACGC GATGCTGCGC AGCAAGCGCC TTCAGGCTCC GGAAATCGCG   840
GCCCTGCTGG AACGGGTTCT GCCCTACGCC GCGCGCGCCT AG                      882
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Arg Arg Asp Val Asn Pro Ile His Ala Thr Leu Leu Gln Thr Arg
1               5                   10                  15

Leu Glu Glu Ile Ala Gln Gly Phe Gly Ala Val Ser Gln Pro Leu Gly
            20                  25                  30

Pro Ala Met Ser His Gly Ala Leu Ser Ser Gly Lys Arg Phe Arg Gly
        35                  40                  45

Met Leu Met Leu Leu Ala Ala Glu Ala Ser Gly Gly Val Cys Asp Thr
```

```
    50                  55                  60

Ile Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu
65                  70                  75                  80

Ile Phe Asp Asp Leu Pro Cys Met Asp Asp Ala Gly Leu Arg Arg Gly
                85                  90                  95

Gln Pro Ala Thr His Val Ala His Gly Glu Ser Arg Ala Val Leu Gly
            100                 105                 110

Gly Ile Ala Leu Ile Thr Glu Ala Met Ala Leu Leu Ala Gly Ala Arg
        115                 120                 125

Gly Ala Ser Gly Thr Val Arg Ala Gln Leu Val Arg Ile Leu Ser Arg
    130                 135                 140

Ser Leu Gly Pro Gln Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His
145                 150                 155                 160

Ala Ala Lys Asn Gly Ala Gly Val Glu Gln Glu Gln Asp Leu Lys Thr
                165                 170                 175

Gly Val Leu Phe Ile Ala Gly Leu Glu Met Leu Ala Val Ile Lys Glu
            180                 185                 190

Phe Asp Ala Glu Glu Gln Thr Gln Met Ile Asp Phe Gly Arg Gln Leu
        195                 200                 205

Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Val Gly Asp
    210                 215                 220

Gln Ala Ala Leu Gly Lys Asp Thr Gly Arg Asp Ala Ala Ala Pro Gly
225                 230                 235                 240

Pro Arg Arg Gly Leu Leu Ala Val Ser Asp Leu Gln Asn Val Ser Arg
                245                 250                 255

His Tyr Glu Ala Ser Arg Ala Gln Leu Asp Ala Met Leu Arg Ser Lys
            260                 265                 270

Arg Leu Gln Ala Pro Glu Ile Ala Ala Leu Leu Glu Arg Val Leu Pro
        275                 280                 285

Tyr Ala Ala Arg Ala
    290
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 295 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Met Thr Pro Lys Gln Gln Phe Pro Leu Arg Asp Leu Val Glu Ile Arg
1               5                   10                  15

Leu Ala Gln Ile Ser Gly Gln Phe Gly Val Val Ser Ala Pro Leu Gly
            20                  25                  30

Ala Ala Met Ser Asp Ala Ala Leu Ser Pro Gly Lys Arg Phe Arg Ala
        35                  40                  45
```

```
Val Leu Met Leu Met Val Ala Glu Ser Ser Gly Gly Val Cys Asp Ala
    50                  55                  60

Met Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu
65                  70                  75                  80

Ile Phe Asp Asp Met Pro Cys Met Asp Asp Ala Arg Thr Arg Arg Gly
                85                  90                  95

Gln Pro Ala Thr His Val Ala His Gly Glu Gly Arg Ala Val Leu Ala
            100                 105                 110

Gly Ile Ala Leu Ile Thr Glu Ala Met Arg Ile Leu Gly Glu Ala Arg
        115                 120                 125

Gly Ala Thr Pro Asp Gln Arg Ala Arg Leu Val Ala Ser Met Ser Arg
    130                 135                 140

Ala Met Gly Pro Val Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His
145                 150                 155                 160

Ala Pro Lys Asp Ala Ala Gly Ile Glu Arg Glu Gln Asp Leu Lys Thr
                165                 170                 175

Gly Val Leu Phe Val Ala Gly Leu Glu Met Leu Ser Ile Ile Lys Gly
            180                 185                 190

Leu Asp Lys Ala Glu Thr Glu Gln Leu Met Ala Phe Gly Arg Gln Leu
        195                 200                 205

Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Ile Gly Asp
    210                 215                 220

Lys Ala Ser Thr Gly Lys Asp Thr Ala Arg Asp Thr Ala Ala Pro Gly
225                 230                 235                 240

Pro Lys Gly Gly Leu Met Ala Val Gly Gln Met Gly Asp Val Ala Gln
                245                 250                 255

His Tyr Arg Ala Ser Arg Ala Gln Leu Asp Glu Leu Met Arg Thr Arg
            260                 265                 270

Leu Phe Arg Gly Gly Gln Ile Ala Asp Leu Leu Ala Arg Val Leu Pro
        275                 280                 285

His Asp Ile Arg Arg Ser Ala
    290                 295
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 888 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
ATGACGCCCA AGCAGCAATT CCCCCTACGC GATCTGGTCG AGATCAGGCT GGCGCAGATC    60
TCGGGCCAGT TCGGCGTGGT CTCGGCCCCG CTCGGCGCGG CCATGAGCGA TGCCGCCCTG   120
TCCCCCGGCA AACGCTTTCG CGCCGTGCTG ATGCTGATGG TCGCCGAAAG CTCGGGCGGG   180
GTCTGCGATG CGATGGTCGA TGCCGCCTGC GCGGTCGAGA TGGTCCATGC CGCATCGCTG   240
ATCTTCGACG ACATGCCCTG CATGGACGAT GCCAGGACCC GTCGCGGTCA GCCCGCCACC   300
```

CATGTCGCCC ATGGCGAGGG GCGCGCGGTG CTTGCGGGCA TCGCCCTGAT CACCGAGGCC 360

ATGCGGATTT TGGGCGAGGC GCGCGGCGCG ACGCCGGATC AGCGCGCAAG GCTGGTCGCA 420

TCCATGTCGC GCGCGATGGG ACCGGTGGGG CTGTGCGCAG GGCAGGATCT GGACCTGCAC 480

GCCCCCAAGG ACGCCGCCGG GATCGAACGT GAACAGGACC TCAAGACCGG CGTGCTGTTC 540

GTCGCGGGCC TCGAGATGCT GTCCATTATT AAGGGTCTGG ACAAGGCCGA GACCGAGCAG 600

CTCATGGCCT TCGGGCGTCA GCTTGGTCGG GTCTTCCAGT CCTATGACGA CCTGCTGGAC 660

GTGATCGGCG ACAAGGCCAG CACCGGCAAG GATACGGCGC GCGACACCGC CGCCCCCGGC 720

CCAAAGGGCC GCCTGATGCC GCTCGGACAG ATGGCCGACG TGGCGCAGCA TTACCGCGCC 780

AGCCGCGCGC AACTGGACGA GCTGATGCGC ACCCGGCTGT TCCGCGGGGG GCAGATCGCG 840

GACCTGCTGG CCCGCGTGCT GCCGCATGAC ATCCGCCGCA GCGCCTAG 888

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
Met Thr Asp Leu Thr Ala Thr Ser Glu Ala Ala Ile Ala Gln Gly Ser
1               5                   10                  15

Gln Ser Phe Ala Gln Ala Ala Lys Leu Met Pro Pro Gly Ile Arg Glu
            20                  25                  30

Asp Thr Val Met Leu Tyr Ala Trp Cys Arg His Ala Asp Asp Val Ile
        35                  40                  45

Asp Gly Gln Val Met Gly Ser Ala Pro Glu Ala Gly Gly Asp Pro Gln
    50                  55                  60

Ala Arg Leu Gly Ala Leu Arg Ala Asp Thr Leu Ala Ala Leu His Glu
65                  70                  75                  80

Asp Gly Pro Met Ser Pro Pro Phe Ala Ala Leu Arg Gln Val Ala Arg
                85                  90                  95

Arg His Asp Phe Pro Asp Leu Trp Pro Met Asp Leu Ile Glu Gly Phe
            100                 105                 110

Ala Met Asp Val Ala Asp Arg Glu Tyr Arg Ser Leu Asp Asp Val Leu
        115                 120                 125

Glu Tyr Ser Tyr His Val Ala Gly Val Val Gly Val Met Met Ala Arg
    130                 135                 140

Val Met Gly Val Gln Asp Asp Ala Val Leu Asp Arg Ala Cys Asp Leu
145                 150                 155                 160

Gly Leu Ala Phe Gln Leu Thr Asn Ile Ala Arg Asp Val Ile Asp Asp
                165                 170                 175

Ala Ala Ile Gly Arg Cys Tyr Leu Pro Ala Asp Trp Leu Ala Glu Ala
            180                 185                 190
```

```
Gly Ala Thr Val Glu Gly Pro Val Pro Ser Asp Ala Leu Tyr Ser Val
        195                 200                 205

Ile Ile Arg Leu Leu Asp Ala Ala Glu Pro Tyr Tyr Ala Ser Ala Arg
    210                 215                 220

Gln Gly Leu Pro His Leu Pro Pro Arg Cys Ala Trp Ser Ile Ala Ala
225                 230                 235                 240

Ala Leu Arg Ile Tyr Arg Ala Ile Gly Thr Arg Ile Arg Gln Gly Gly
                245                 250                 255

Pro Glu Ala Tyr Arg Gln Arg Ile Ser Thr Ser Lys Ala Ala Lys Ile
            260                 265                 270

Gly Leu Leu Ala Arg Gly Gly Leu Asp Ala Ala Ala Ser Arg Leu Arg
        275                 280                 285

Gly Gly Glu Ile Ser Arg Asp Gly Leu Trp Thr Arg Pro Arg Ala
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
ATGACCGATC TGACGGCGAC TTCCGAAGCG GCCATCGCGC AGGGTTCGCA AAGCTTCGCG    60
CAGGCGGCCA AGCTGATGCC GCCCGGCATC CGCGAGGATA CGGTCATGCT CTATGCCTGG   120
TGCAGGCATG CGGATGACGT GATCGACGGG CAGGTGATGG GTTCTGCCCC CGAGGCGGGC   180
GGCGACCCAC AGGCGCGGCT GGGGGCGCTG CGCGCCGACA CGCTGGCCGC GCTGCACGAG   240
GACGGCCCGA TGTCGCCGCC CTTCGCGGCG CTGCGCCAGG TCGCCCGGCG GCATGATTTC   300
CCGGACCTTT GGCCGATGGA CCTGATCGAG GGTTTCGCGA TGGATGTCGC GGATCGCGAA   360
TACCGCAGCC TGGATGACGT GCTGGAATAT TCCTACCACG TCGCGGGGGT CGTGGGCGTG   420
ATGATGGCGC GGGTGATGGG CGTGCAGGAC GATGCGGTGC TGGATCGCGC CTGCGATCTG   480
GGCCTTGCGT TCCAGCTGAC GAACATCGCT CGCGACGTGA TCGACGATGC CGCCATCGGG   540
CGCTGCTATC TGCCTGCCGA CTGGCTGGCC GAGGCGGGGG CGACGGTTGA GGGTCCGGTG   600
CCTTCGGACG CGCTCTATTC CGTCATCATC CGCCTGCTTG ACGCGGCCGA GCCCTATTAT   660
GCCTCGGCGC GGCAGGGGCT TCCGCATCTG CCGCCGCGCT GCGCGTGGTC GATCGCCGCC   720
GCGCTGCGTA TCTATCGCGC AATCGGGACG CGCATCCGGC AGGGTGGCCC CGAGGCCTAT   780
CGCCAGCGGA TCAGCACGTC GAAGGCTGCC AAGATCGGGC TTCTGGCGCG CGGAGGCTTG   840
GACGCGGCCG CATCGCGCCT GCGCGGCGGC GAAATCAGCC GCGACGGCCT GTGGACCCGA   900
CCGCGCGC                                                            908
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
Met Ser Ser Ala Ile Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
1               5                   10                  15

Ala Ile Arg Leu Gln Ser Ala Gly Ile Ala Thr Thr Ile Val Glu Ala
            20                  25                  30

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Trp Asn Asp Gln Gly His
        35                  40                  45

Val Phe Asp Ala Gly Pro Thr Val Val Thr Asp Pro Asp Ser Leu Arg
    50                  55                  60

Glu Leu Trp Ala Leu Ser Gly Gln Pro Met Glu Arg Asp Val Thr Leu
65                  70                  75                  80

Leu Pro Val Ser Pro Phe Tyr Arg Leu Thr Trp Ala Asp Gly Arg Ser
                85                  90                  95

Phe Glu Tyr Val Asn Asp Asp Asp Glu Leu Ile Arg Gln Val Ala Ser
            100                 105                 110

Phe Asn Pro Ala Asp Val Asp Gly Tyr Arg Arg Phe His Asp Tyr Ala
        115                 120                 125

Glu Glu Val Tyr Arg Glu Gly Tyr Leu Lys Leu Gly Thr Thr Pro Phe
    130                 135                 140

Leu Lys Leu Gly Gln Met Leu Asn Ala Ala Pro Ala Leu Met Arg Leu
145                 150                 155                 160

Gln Ala Tyr Arg Ser Val His Ser Met Val Ala Arg Phe Ile Gln Asp
                165                 170                 175

Pro His Leu Arg Gln Ala Phe Ser Phe His Thr Leu Leu Val Gly Gly
            180                 185                 190

Asn Pro Phe Ser Thr Ser Ser Ile Tyr Ala Leu Ile His Ala Leu Glu
        195                 200                 205

Arg Arg Gly Gly Val Trp Phe Ala Lys Gly Gly Thr Asn Gln Leu Val
    210                 215                 220

Ala Gly Met Val Ala Leu Phe Thr Arg Leu Trp Trp Thr Leu Leu Leu
225                 230                 235                 240

Asn Ala Arg Val Thr Arg Ile Asp Thr Glu Gly Asp Arg Ala Thr Gly
                245                 250                 255

Val Thr Leu Leu Asp Gly Arg Gln Leu Arg Ala Asp Thr Val Ala Ser
            260                 265                 270

Asn Gly Asp Val Met His Ser Tyr Arg Asp Leu Leu Gly His Thr Arg
        275                 280                 285

Arg Gly Arg Thr Lys Ala Ala Ile Leu Asn Arg Gln Arg Trp Ser Met
    290                 295                 300
```

```
Ser Leu Phe Val Leu His Phe Gly Leu Ser Lys Arg Pro Glu Asn Leu
305                 310                 315                 320

Ala His His Ser Val Ile Phe Gly Pro Arg Tyr Lys Gly Leu Val Asn
                325                 330                 335

Glu Ile Phe Asn Gly Pro Arg Leu Pro Asp Asp Phe Ser Met Tyr Leu
            340                 345                 350

His Ser Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Glu Gly Met Ser
        355                 360                 365

Thr His Tyr Val Leu Ala Pro Val Pro His Leu Gly Arg Ala Asp Val
    370                 375                 380

Asp Trp Glu Ala Glu Ala Pro Gly Tyr Ala Glu Arg Ile Phe Glu Glu
385                 390                 395                 400

Leu Glu Arg Arg Ala Ile Pro Asp Leu Arg Lys His Leu Thr Val Ser
                405                 410                 415

Arg Ile Phe Ser Pro Ala Asp Phe Ser Thr Glu Leu Ser Ala His His
            420                 425                 430

Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe
        435                 440                 445

Arg Pro His Asn Arg Asp Arg Ala Ile Pro Asn Phe Tyr Ile Val Gly
    450                 455                 460

Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Val Gly Ser Ala
465                 470                 475                 480

Lys Ala Thr Ala Gln Val Met Leu Ser Asp Leu Ala Val Ala
                485                 490
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
ATGAGTTCCG CCATCGTCAT CGGCGCAGGT TTCGGCGGGC TTGCGCTTGC CATCCGCCTG    60
CAATCGGCCG GCATCGCGAC CACCATCGTC GAGGCCCGCG ACAAGCCCGG CGGCCGCGCC   120
TATGTCTGGA ACGATCAGGG CCACGTCTTC GATGCAGGCC CGACGGTCGT GACCGACCCC   180
GACAGCCTGC GAGAGCTGTG GGCCCTCAGC GGCCAACCGA TGGAGCGTGA CGTGACGCTG   240
CTGCCGGTCT CGCCCTTCTA CCGGCTGACA TGGGCGGACG GCCGCAGCTT CGAATACGTG   300
AACGACGACG ACGAGCTGAT CCGCCAGGTC GCCTCCTTCA ATCCCGCCGA TGTCGATGGC   360
TATCGCCGCT TCCACGATTA CGCCGAGGAG GTCTATCGCG AGGGGTATCT GAAGCTGGGG   420
ACCACGCCCT TCCTGAAGCT GGGCCAGATG CTGAACGCCG CGCCGGCGCT GATGCGCCTG   480
CAGGCATACC GCTCGGTCCA CAGCATGGTG GCGCGCTTCA TCCAGGACCC GCATCTGCGG   540
CAGGCCTTCT CGTTCCACAC GCTGCTGGTC GGCGGGAACC CGTTTTCGAC CAGCTCGATC   600
TATGCGCTGA TCCATGCGCT GGAACGGCGC GGCGGCGTCT GGTTCGCCAA GGGCGGCACC   660
```

```
AACCAGCTGG TCGCGGGCAT GGTCGCCCTG TTCGAGCGTC TTGGCGGCAC GCTGCTGCTG    720
AATGCCCGCG TCACGCGGAT CGACACCGAG GGCGATCGCG CCACGGGCGT CACGCTGCTG    780
GACGGGCGGC AGTTGCGCGC GGATACGGTG GCCAGCAACG GCGACGTGAT GCACAGCTAT    840
CGCGACCTGC TGGGCCATAC CCGCCGCGGG CGCACCAAGG CCGCGATCCT GAACCGGCAG    900
CGCTGGTCGA TGTCGCTGTT CGTGCTGCAT TTCGGCCTGT CCAAGCGCCC CGAGAACCTG    960
GCCCACCACA GCGTCATCTT CGGCCCGCGC TACAAGGGGC TGGTGAACGA GATCTTCAAC   1020
GGGCCACGCC TGCCGGACGA TTTCTCGATG TATCTGCATT CGCCCTGCGT GACCGATCCC   1080
AGCCTGGCCC CCGAGGGGAT GTCCACGCAT TACGTCCTTG CGCCCGTTCC GCATCTGGGC   1140
CGCGCCGATG TCGATTGGGA AGCCGAGGCC CCGGGCTATG CCGAGCGCAT CTTCGAGGAA   1200
CTGGAGCGCC GCGCCATCCC CGACCTGCGC AAGCACCTGA CCGTCAGCCG CATCTTCAGC   1260
CCCGCCGATT TCAGCACCGA ACTGTCGGCC CATCACGGCA GCGCCTTCTC GGTCGAGCCG   1320
ATCCTGACGC AATCCGCCTG GTTCCGCCCG CATAACCGCG ACCGCGCGAT CCCGAACTTC   1380
TACATCGTGG GGGCGGGCAC GCATCCGGGT GCGGGCATCC CGGGTGTCGT TGGCAGCGCC   1440
AAGGCCACGG CGCAGGTCAT GCTGTCGGAC CTGGCCGTCG CA                       1482
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
Met Ser His Asp Leu Leu Ile Ala Gly Ala Gly Leu Ser Gly Ala Leu
1               5                   10                  15

Ile Ala Leu Ala Val Arg Asp Arg Arg Pro Asp Ala Arg Ile Val Met
            20                  25                  30

Leu Asp Ala Arg Ser Gly Pro Ser Asp Gln His Thr Trp Ser Cys His
        35                  40                  45

Asp Thr Asp Leu Ser Pro Glu Trp Leu Ala Arg Leu Ser Pro Ile Arg
    50                  55                  60

Arg Gly Glu Trp Thr Asp Gln Glu Val Ala Phe Pro Asp His Ser Arg
65                  70                  75                  80

Arg Leu Thr Thr Gly Tyr Gly Ser Ile Glu Ala Gly Ala Leu Ile Gly
                85                  90                  95

Leu Leu Gln Gly Val Asp Leu Arg Trp Asn Thr His Val Ala Thr Leu
            100                 105                 110

Asp Asp Thr Gly Ala Thr Leu Thr Asp Gly Ser Arg Ile Glu Ala Ala
        115                 120                 125

Cys Val Ile Asp Ala Arg Gly Ala Val Glu Thr Pro His Leu Thr Val
    130                 135                 140

Gly Phe Gln Lys Phe Val Gly Val Glu Ile Glu Thr Asp Ala Pro His
```

```
145                 150                 155                 160

Gly Val Glu Arg Pro Met Ile Met Asp Ala Thr Val Pro Gln Met Asp
                165                 170                 175

Gly Tyr Arg Phe Ile Tyr Leu Leu Pro Phe Ser Pro Thr Arg Ile Leu
            180                 185                 190

Ile Glu Asp Thr Arg Tyr Ser Asp Gly Gly Asp Leu Asp Asp Gly Ala
        195                 200                 205

Leu Ala Gln Ala Ser Leu Asp Tyr Ala Ala Arg Arg Gly Trp Thr Gly
    210                 215                 220

Gln Glu Met Arg Arg Glu Arg Gly Ile Leu Pro Ile Ala Leu Ala His
225                 230                 235                 240

Asp Ala Ile Gly Phe Trp Arg Asp His Ala Gln Gly Ala Val Pro Val
                245                 250                 255

Gly Leu Gly Ala Gly Leu Phe His Pro Val Thr Gly Tyr Ser Leu Pro
            260                 265                 270

Tyr Ala Ala Gln Val Ala Asp Ala Ile Ala Ala Arg Asp Leu Thr Thr
        275                 280                 285

Ala Ser Ala Arg Arg Ala Val Arg Gly Trp Ala Ile Asp Arg Ala Asp
    290                 295                 300

Arg Asp Arg Phe Leu Arg Leu Leu Asn Arg Met Leu Phe Arg Gly Cys
305                 310                 315                 320

Pro Pro Asp Arg Arg Tyr Arg Leu Leu Gln Arg Phe Tyr Arg Leu Pro
                325                 330                 335

Gln Pro Leu Ile Glu Arg Phe Tyr Ala Gly Arg Leu Thr Leu Ala Asp
            340                 345                 350

Arg Leu Arg Ile Val Thr Gly Arg Pro Pro Ile Pro Leu Ser Gln Ala
        355                 360                 365

Val Arg Cys Leu Pro Glu Arg Pro Leu Leu Gln Glu Arg Ala
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1149 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
ATGAGCCATG ATCTGCTGAT CGCGGGCGCG GGGCTGTCCG GTGCGCTGAT CGCGCTTGCC     60
GTTCGCGACC GCAGACCGGA TGCGCGCATC GTGATGCTCG ACGCGCGGTC CGGCCCCTCG    120
GACCAGCACA CCTGGTCCTG CCACGACACG GATCTTTCGC CCGAATGGCT GGCGCGCCTG    180
TCGCCCATTC GTCGCGGCGA ATGGACGGAT CAGGAGGTCG CGTTTCCCGA CCATTCGCGC    240
CGCCTGACGA CAGGCTATGG CTCGATCGAG GCGGGCGCGC TGATCGGGCT GCTGCAGGGT    300
GTCGATCTGC GGTGGAATAC GCATGTCGCG ACGCTGGACG ATACCGGCGC GACGCTGACG    360
GACGGCTCGC GGATCGAGGC TGCCTGCGTG ATCGACGCCC GTGGTGCCGT CGAGACCCCG    420
```

CACCTGACCG TGGGTTTCCA GAAATTCGTG GGCGTCGAGA TCGAGACCGA CGCCCCCCAT 480
GGCGTCGAGC GCCCGATGAT CATGGACGCG ACCGTTCCGC AGATGGACGG GTACCGCTTC 540
ATCTATCTGC TGCCCTTCAG TCCCACCCGC ATCCTGATCG AGGATACGCG CTACAGCGAC 600
GGCGGCGATC TGGACGATGG CGCGCTGGCG CAGGCGTCGC TGGACTATGC CGCCAGGCGG 660
GGCTGGACCG GGCAGGAGAT GCGGCGCGAA AGGGGCATCC TGCCCATCGC GCTGGCCCAT 720
GACGCCATAG GCTTCTGGCG CGACCACGCG CAGGGGGCGG TGCCGGTTGG GCTGGGGGCA 780
GGGCTGTTCC ACCCCGTCAC CGGATATTCG CTGCCCTATG CCGCGCAGGT CGCGGATGCC 840
ATCGCGGCGC GCGACCTGAC GACCGCGTCC GCCCGTCGCG CGGTGCGCGG CTGGGCCATC 900
GATCGCGCGG ATCGCGACCG CTTCCTGCGG CTGCTGAACC GGATGCTGTT CCGCGGCTGC 960
CCGCCCGACC GTCGCTATCG CCTGCTGCAG CGGTTCTACC GCCTGCCGCA GCCGCTGATC 1020
GAGCGCTTCT ATGCCGGGCG CCTGACATTG GCCGACCGGC TTCGCATCGT CACCGGACGC 1080
CCGCCCATTC CGCTGTCGCA GGCCGTGCGC TGCCTGCCCG AACGCCCCCT GCTGCAGGAG 1140
AGAGCATGA 1149

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
Met Ser Thr Trp Ala Ala Ile Leu Thr Val Ile Leu Thr Val Ala Ala
1               5                   10                  15

Met Glu Leu Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro
            20                  25                  30

Leu Gly Trp Gly Trp His Lys Ser His His Asp Glu Asp His Asp His
        35                  40                  45

Ala Leu Glu Lys Asn Asp Leu Tyr Gly Val Ile Phe Ala Val Ile Ser
    50                  55                  60

Ile Val Leu Phe Ala Ile Gly Ala Met Gly Ser Asp Leu Ala Trp Trp
65                  70                  75                  80

Leu Ala Val Gly Val Thr Cys Tyr Gly Leu Ile Tyr Tyr Phe Leu His
                85                  90                  95

Asp Gly Leu Val His Gly Arg Trp Pro Phe Arg Tyr Val Pro Lys Arg
            100                 105                 110

Gly Tyr Leu Arg Arg Val Tyr Gln Ala His Arg Met His His Ala Val
        115                 120                 125

His Gly Arg Glu Asn Cys Val Ser Phe Gly Phe Ile Trp Ala Pro Ser
    130                 135                 140

Val Asp Ser Leu Lys Ala Glu Leu Lys Arg Ser Gly Ala Leu Leu Lys
145                 150                 155                 160
```

Asp Arg Glu Gly Ala Asp Arg Asn Thr
165

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 506 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
ATGAGCACTT GGGCCGCAAT CCTGACCGTC ATCCTGACCG TCGCCGCGAT GGAGCTGACG    60
GCCTACTCCG TCCATCGGTG GATCATGCAT GGCCCCCTGG GCTGGGGCTG GCATAAATCG   120
CACCACGACG AGGATCACGA CCACGCGCTC GAGAAGAACG ACCTCTATGG CGTCATCTTC   180
GCGGTAATCT CGATCGTGCT GTTCGCGATC GGCGCGATGG GGTCGGATCT GGCCTGGTGG   240
CTGGCGGTGG GGGTCACCTG CTACGGGCTG ATCTACTATT TCCTGCATGA CGGCTTGGTG   300
CATGGGCGCT GGCCGTTCCG CTATGTCCCC AAGCGCGGCT ATCTTCGTCG CGTCTACCAG   360
GCACACAGGA TGCATCACGC GGTCCATGGC CGCGAGAACT GCGTCAGCTT CGGTTTCATC   420
TGGGCGCCCT CGGTCGACAG CCTCAAGGCA GAGCTGAAAC GCTCGGGCGC GCTGCTGAAG   480
GACCGCGAAG GGGCGGATCG CAATAC                                        506
```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
ATGTCCGGTC GTAAACCGGG TACCACCGGT GACACCATCG TTAACCTGGG TCTGACCGCT    60
GCTATCCTGC TGTGCTGGCT GGTTCTGCAC GCTTTCACCC TGTGGCTGCT GGACGCTGCT   120
GCTCACCCGC TGCTGGCTGT TCTGTGCCTG GCTGGTCTGA CCTGGCTGTC CGTTGGTCTG   180
TTCATCATCG CTCACGACGC TATGCACGGT TCCGTTGTTC CGGGTCGTCC GCGGGCTAAC   240
GCTGCTATCG GTCAGCTGGC TCTGTGGCTG TACGCTGGTT TCTCCTGGCC GAAACTGATC   300
GCTAAACACA TGACCCACCA CCGTCACGCT GGTACCGACA ACGACCCGGA CTTCGGTCAC   360
GGTGGTCCGG TTCGTTGGTA CGGTTCCTTC GTTTCCACCT ACTTCGGTTG GCGTGAAGGT   420
CTGCTGCTGC CGGTTATCGT TACCACCTAC GCTCTGATCC TGGGTGACCG TTGGATGTAC   480
GTTATCTTCT GGCCGGTTCC GGCTGTTCTG GCTTCCATCC AGATCTTCGT TTTCGGTACC   540
TGGCTGCCGC ACCGTCCGGG TCACGACGAC TTCCCGGACC GTCACAACGC TCGTTCCACC   600
GGTATCGGTG ACCCGCTGTC CCTGCTGACC TGCTTCCACT TCGGTGGTTA CCACCACGAA   660
```

CACCACCTGC ACCCGCACGT TCCGTGGTGG CGTCTGCCGC GTACCCGTAA AACCGGTGGT 720

CGTGCT 726

Claims

1. A process for the preparation of canthaxanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following DNA sequences:

   a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

   b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

   c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtI) or a DNA sequence which is substantially homologous;

   d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous;

   e) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E-396 (FERM BP-4283) [$crtW_{E396}$] or a DNA sequence which is substantially homologous;

   or a cell which is transformed by a vector comprising DNA sequences specified above under a) to e) and by isolating canthaxanthin from such cells or the culture medium by methods known in the art.

2. A process for the preparation of a mixture of adonixanthin and astaxanthin or adonixanthin or astaxanthin alone by a process as claimed in claim 1 characterized therein that in addition to the DNA sequences specified in claim 1 under a) to e) the following additional DNA sequence is present:

   f) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E-396 (FERM BP-4283) [$crtZ_{E396}$] or a DNA sequence which is substantially homologous;

   and the DNA sequence specified under e) of claim 1 is as specified in claim 1 or the following sequence:

   g) a DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crtW) or a DNA sequence which is substantially homologous;

   and isolating the desired mixture of adonixanthin and astaxanthin or adonixanthin or a astaxanthin alone from such cells of the culture medium and separating the desired mixture or carotenoids alone from other carotenoids which might be present by methods known in the art.

3. A process for the preparation of zeaxanthin by a process as claimed in claim 1 characterized therein that the DNA sequence as specified under e) is replaced by the DNA sequence as specified under f) in claim 2 and by isolating zeaxanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

4. A process for the production of adonixanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following heterologous DNA sequences:

   a) a DNA sequence which encodes the GGPP synthase of the microorganism E-396 (FERM BP-4283)

[$crtE_{E396}$] or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of the microorganism E-396 (FERM BP-4283) [$crtB_{E396}$] or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of the microorganism E-396 (FERM BP-4283) [$crtI_{E396}$] or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of the microorganism E-396 (FERM BP-4283) [$crtY_{E396}$] or a DNA sequence which is substantially homologous;

e) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E396 (FERM BP-4283) [$crtZ_{E396}$] or a DNA sequence which is substantially homologous; and

f) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E396 (FERM BP-4283) [$crtW_{E396}$] or a DNA sequence which is substantially homologous;

and isolating adonixanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

5. A process for the preparation of a food or feed composition characterized therein that after a process as claimed in any one of claims 1 to 4 has been effected the carotenoid or carotenoid mixture is added to food or feed.

6. A process as claimed in any one of claims 1 to 5 characterized therein that the transformed host cell is a prokaryotic host cell, like E. coli, Bacillus or Flavobacter.

7. A process as claimed in any one of claims 1 to 5 characterized therein that the transformed host cell is a eukaryotice host cell, like yeast or a fungal cell.

Fig. 1

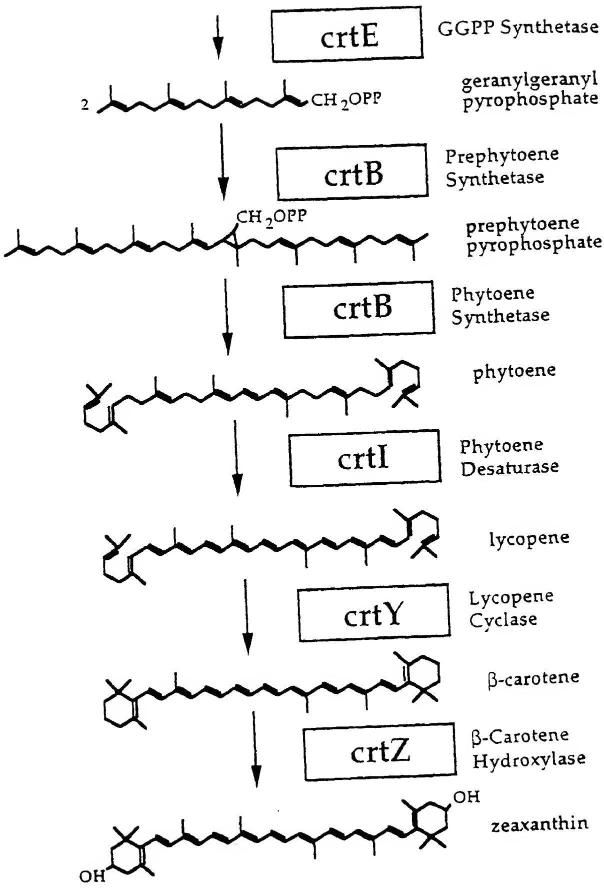

crtE
GGPP Synthetase
2
CH2OPP
geranylgeranyl pyrophosphate
crtB
Prephytoene Synthetase
CH2OPP
prephytoene pyrophosphate
crtB
Phytoene Synthetase
phytoene
crtI
Phytoene Desaturase
lycopene
crtY
Lycopene Cyclase
β-carotene
crtZ
β-Carotene Hydroxylase
OH
zeaxanthin
OH

Fig. 2

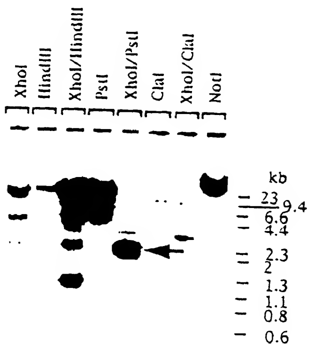

XhoI
HindIII
XhoI/HindIII
PstI
XhoI/PstI
ClaI
XhoI/ClaI
NotI
kb
23
9.4
6.6
4.4
2.3
2
1.3
1.1
0.8
0.6

Fig. 3

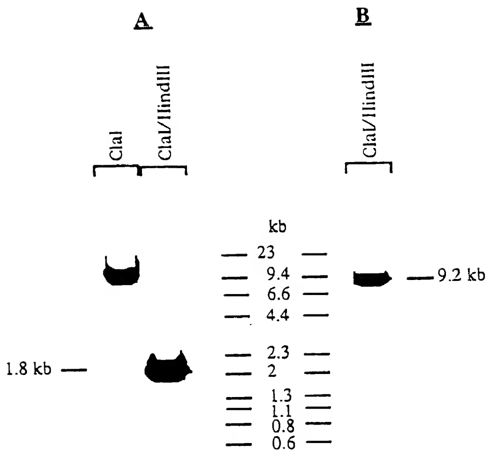

A
B
ClaI
ClaI/HindIII
ClaI/HindIII
kb
23
9.4
6.6
4.4
2.3
2
1.3
1.1
0.8
0.6
1.8 kb
9.2 kb

Fig. 4

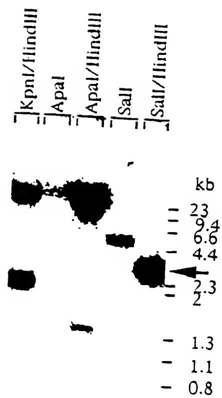

KpnI/HindIII
ApaI
ApaI/HindIII
SalI
SalI/HindIII
kb
23
9.4
6.6
4.4
2.3
2
1.3
1.1
0.8

Fig. 5

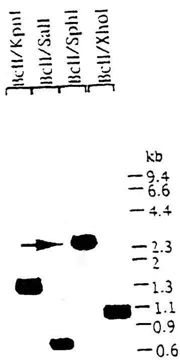

BclI/KpnI
BclI/SalI
BclI/SphI
BclI/XhoI
kb
9.4
6.6
4.4
2.3
2
1.3
1.1
0.9
0.6

Fig. 6

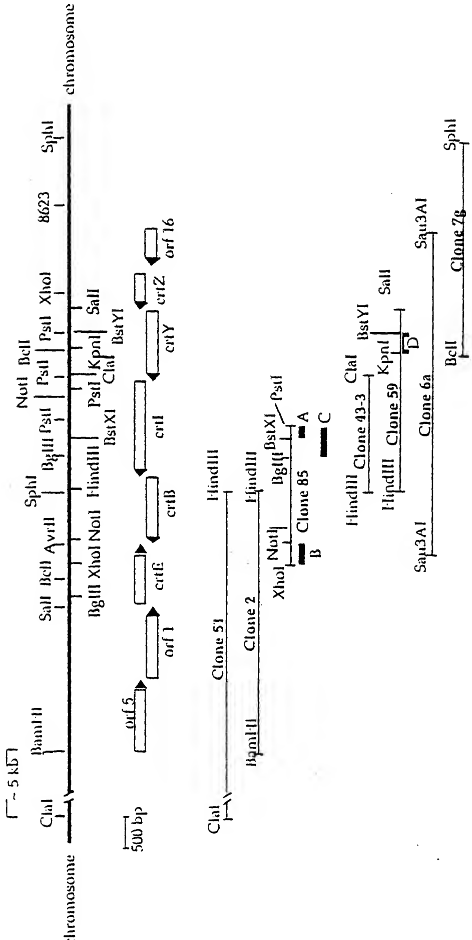

chromosome
ClaI
~5 kb
BamHI
SalI
BclI
BglII
XhoI
AvrII
NotI
SphI
HindIII
PstI
BstXI
KpnI
ClaI
BstYI
8623
500 bp
orf 5
orf 1
crtE
crtB
crtI
crtY
crtZ
orf 16
Clone 51
Clone 2
Clone 85
A
B
C
Clone 43-3
Clone 59
D
Sau3AI
Clone 6a
Clone 7b
chromosome

Fig. 7/1

```
        GGATCCGCGCCTGGCCGTTCGCGATCAGCAGCCGCCCTTGCGGATCGGTC
1       ---------+---------+---------+---------+---------+ 50
        CCTAGGCGCGGACCGGCAAGCGCTAGTCGTCGGCGGGAACGCCTAGCCAG

orf-5 --> D  P  R  L  A  V  R  D  Q  Q  P  P  L  R  I  G  Q

        AGCATCATCCCCATGAACCGCAGCGCACGACGCAGCGCGCGCCCCAGATC
51      ---------+---------+---------+---------+---------+ 100
        TCGTAGTAGGGGTACTTGGCGTCGCGTGCTGCGTCGCGCGCGGGGTCTAG

          H  H  P  H  E  P  Q  R  T  T  Q  R  A  P  Q  I

        GGGCGCGTCCAGCACGGCATGCGCCATCATCGCGAAGGCCCCCGGCGGCA
101     ---------+---------+---------+---------+---------+ 150
        CCCGCGCAGGTCGTGCCGTACGCGGTAGTAGCGCTTCCGGGGGCCGCCGT

        G  R  V  Q  H  G  M  R  H  H  R  K  G  P  R  R  H

        TGGGGCGCGTGCCCATTCCGAAGAACTCGCAGCCTGTCCGCTGCGCAAGG
151     ---------+---------+---------+---------+---------+ 200
        ACCCCGCGCACGGGTAAGGCTTCTTGAGCGTCGGACAGGCGACGCGTTCC

         G  A  R  A  H  S  E  E  L  A  A  C  P  L  R  K  V

        TCGCGCCAGATCGCGCCGTATTCCGATGCAGTGACGGGCCCGATGCGCGT
201     ---------+---------+---------+---------+---------+ 250
        AGCGCGGTCTAGCGCGGCATAAGGCTACGTCACTGCCCGGGCTACGCGCA

          A  P  D  R  A  V  F  R  C  S  D  G  P  D  A  R

        GGGCCCGCCCTGCCCCGCCGCCACCAGCGCATCGCGCACGAACCCTTCCG
251     ---------+---------+---------+---------+---------+ 300
        CCCGGGCGGGACGGGGCGGCGGTGGTCGCGTAGCGCGTGCTTGGGAAGGC

        G  P  A  L  P  R  R  H  Q  R  I  A  H  E  P  F  R

        AGATGATGTGCTGATCCATGGCCCGTCATTGCAAAACCGATCACCGATCC
301     ---------+---------+---------+---------+---------+ 350
        TCTACTACACGACTAGGTACCGGGCAGTAACGTTTTGGCTAGTGGCTAGG

         D  D  V  L  I  H  G  P  S  L  Q  N  R  S  P  I  L

        TGTCGCGTGATGGCATTGTTTGCAATGCCCCGAGGGCTAGGATGGCGCGA
351     ---------+---------+---------+---------+---------+ 400
        ACAGCGCACTACCGTAACAAACGTTACGGGGCTCCCGATCCTACCGCGCT

          S  R  D  G  I  V  C  N  A  P  R  A  R  M  A  R

        AGGATCAAGGGGGGGAGAGACATGGAAATCGAGGGACGGGTCTTTGTCGT
401     ---------+---------+---------+---------+---------+ 450
        TCCTAGTTCCCCCCCTCTCTGTACCTTTAGCTCCCTGCCCAGAAACAGCA

        R  I  K  G  G  R  D  M  E  I  E  G  R  V  F  V  V

        CACGGGCGCCGCATCGGGTCTGGGGGCGGCCTCGGCGCGGATGCTGGCCC
451     ---------+---------+---------+---------+---------+ 500
        GTGCCCGCGGCGTAGCCCAGACCCCCGCCGGAGCCGCGCCTACGACCGGG

         T  G  A  A  S  G  L  G  A  A  S  A  R  M  L  A  Q

        AAGGCGGCGCGAAGGTCGTGCTGGCCGATCTGGCGGAACCGAAGGACGCG
501     ---------+---------+---------+---------+---------+ 550
        TTCCGCCGCGCTTCCAGCACGACCGGCTAGACCGCCTTGGCTTCCTGCGC

          G  G  A  K  V  V  L  A  D  L  A  E  P  K  D  A

        CCCGAAGGCGCGGTTCACGCGGCCTGCGACGTGACCGACGCGACCGCTGC
551     ---------+---------+---------+---------+---------+ 600
        GGGCTTCCGCGCCAAGTGCGCCGGACGCTGCACTGGCTGCGCTGGCGACG

        P  E  G  A  V  H  A  A  C  D  V  T  D  A  T  A  A
```

Fig. 7/2

```
     GCAGACGGCCATCGCGCTGGCGACCGACCGCTTCGGCAGGCTGGACGGCC
 601 ---------+---------+---------+---------+---------+  650
     CGTCTGCCGGTAGCGCGACCGCTGGCTGGCGAAGCCGTCCGACCTGCCGG

      Q  T  A  I  A  L  A  T  D  R  F  G  R  L  D  G  L

     TTGTGAACTGCGCGGGCATCGCGCCGGCCGAACGGATGCTGGGCCGCGAC
 651 ---------+---------+---------+---------+---------+  700
     AACACTTGACGCGCCCGTAGCGCGGCCGGCTTGCCTACGACCCGGCGCTG

       V  N  C  A  G  I  A  P  A  E  R  M  L  G  R  D

     GGGCCGCATGGACTGGACAGCTTTGCCCGTGCGGTCACGATCAACCTGAT
 701 ---------+---------+---------+---------+---------+  750
     CCCGGCGTACCTGACCTGTCGAAACGGGCACGCCAGTGCTAGTTGGACTA

     G  P  H  G  L  D  S  F  A  R  A  V  T  I  N  L  I

     CGGCAGCTTCAACATGGCCCGCCTTGCAGCCGAGGCGATGGCCCGGAACG
 751 ---------+---------+---------+---------+---------+  800
     GCCGTCGAAGTTGTACCGGGCGGAACGTCGGCTCCGCTACCGGGCCTTGC

      G  S  F  N  M  A  R  L  A  A  E  A  M  A  R  N  E

     AGCCCGTCCGGGGCGAGCGTGGCGTGATCGTCAACACGGCCTCGATCGCG
 801 ---------+---------+---------+---------+---------+  850
     TCGGGCAGGCCCCGCTCGCACCGCACTAGCAGTTGTGCCGGAGCTAGCGC

       P  V  R  G  E  R  G  V  I  V  N  T  A  S  I  A

     GCGCAGGACGGACAGATCGGACAGGTCGCCTATGCGGCCAGCAAGGCGGG
 851 ---------+---------+---------+---------+---------+  900
     CGCGTCCTGCCTGTCTAGCCTGTCCAGCGGATACGCCGGTCGTTCCGCCC

     A  Q  D  G  Q  I  G  Q  V  A  Y  A  A  S  K  A  G

     CGTGGCGGGCATGACGCTGCCGATGGCCCGCGACCTTGCGCGGCACGGCA
 901 ---------+---------+---------+---------+---------+  950
     GCACCGCCCGTACTGCGACGGCTACCGGGCGCTGGAACGCGCCGTGCCGT

      V  A  G  M  T  L  P  M  A  R  D  L  A  R  H  G  I

     TCCGCGTCATGACCATCGCGCCCGGCATCTTCCGCACCCCGATGCTGGAG
 951 ---------+---------+---------+---------+---------+  1000
     AGGCGCAGTACTGGTAGCGCGGGCCGTAGAAGGCGTGGGGCTACGACCTC

       R  V  M  T  I  A  P  G  I  F  R  T  P  M  L  E

     GGGCTGCCGCAGGACGTTCAGGACAGCCTGGGCGCGGCGGTGCCCTTCCC
1001 ---------+---------+---------+---------+---------+  1050
     CCCGACGGCGTCCTGCAAGTCCTGTCGGACCCGCGCCGCCACGGGAAGGG

     G  L  P  Q  D  V  Q  D  S  L  G  A  A  V  P  F  P

     CTCGCGGCTGGGAGAGCCGTCGGAATACGCGGCGCTGTTGCACCACATCA
1051 ---------+---------+---------+---------+---------+  1100
     GAGCGCCGACCCTCTCGGCAGCCTTATGCGCCGCGACAACGTGGTGTAGT

      S  R  L  G  E  P  S  E  Y  A  A  L  L  H  H  I  I

     TCGCGAACCCCATGCTGAACGGAGAGGTCATCCGCCTCGACGGCGCATTG
1101 ---------+---------+---------+---------+---------+  1150
     AGCGCTTGGGGTACGACTTGCCTCTCCAGTAGGCGGAGCTGCCGCGTAAC

       A  N  P  M  L  N  G  E  V  I  R  L  D  G  A  L

     CGCATGGCCCCCAAGTGAAGGAGCGTTTCATGGACCCCATCGTCATCACC
1151 ---------+---------+---------+---------+---------+  1200
     GCGTACCGGGGGTTCACTTCCTCGCAAAGTACCTGGGGTAGCAGTAGTGG

     R  M  A  P  K  *                M  D  P  I  V  I  T
                                     orf-1 -->
```

Fig. 7/3

```
     GGCGCGATGCGCACCCCGATGGGGGCATTCCAGGGCGATCTTGCCGCGAT
1201 ---------+---------+---------+---------+---------+ 1250
     CCGCGCTACGCGTGGGGCTACCCCCGTAAGGTCCCGCTAGAACGGCGCTA

     G  A  M  R  T  P  M  G  A  F  Q  G  D  L  A  A  N

     GGATGCCCCGACCCTTGGCGCGGACGCGATCCGCGCCGCGCTGAACGGCC
1251 ---------+---------+---------+---------+---------+ 1300
     CCTACGGGGCTGGGAACCGCGCCTGCGCTAGGCGCGGCGCGACTTGCCGG

      D  A  P  T  L  G  A  D  A  I  R  A  A  L  N  G  L

     TGTCGCCCGACATGGTGGACGAGGTGCTGATGGGCTGCGTCCTCGCCGCG
1301 ---------+---------+---------+---------+---------+ 1350
     ACAGCGGGCTGTACCACCTGCTCCACGACTACCCGACGCAGGAGCGGCGC

       S  P  D  M  V  D  E  V  L  M  G  C  V  L  A  A

     GGCCAGGGTCAGGCACCGGCACGTCAGGCGGCGCTTGGCGCCGGACTGCC
1351 ---------+---------+---------+---------+---------+ 1400
     CCGGTCCCAGTCCGTGGCCGTGCAGTCCGCCGCGAACCGCGGCCTGACGG

     G  Q  G  Q  A  P  A  R  Q  A  A  L  G  A  G  L  P

     GCTGTCGACGGGCACGACCACCATCAACGAGATGTGCGGATCGGGCATGA
1401 ---------+---------+---------+---------+---------+ 1450
     CGACAGCTGCCCGTGCTGGTGGTAGTTGCTCTACACGCCTAGCCCGTACT

      L  S  T  G  T  T  T  I  N  E  M  C  G  S  G  M  K

     AGGCCGCGATGCTGGGCCATGACCTGATCGCCGCGGGATCGGCGGGCATC
1451 ---------+---------+---------+---------+---------+ 1500
     TCCGGCGCTACGACCCGGTACTGGACTAGCGGCGCCCTAGCCGCCCGTAG

       A  A  M  L  G  H  D  L  I  A  A  G  S  A  G  I

     GTCGTCGCCGGCGGGATGGAGAGCATGTCGAACGCCCCCTACCTGCTGCC
1501 ---------+---------+---------+---------+---------+ 1550
     CAGCAGCGGCCGCCCTACCTCTCGTACAGCTTGCGGGGGATGGACGACGG

     V  V  A  G  G  M  E  S  M  S  N  A  P  Y  L  L  P

     CAAGGCGCGGTCGGGGATGCGCATGGGCCATGACCGTGTGCTGGATCACA
1551 ---------+---------+---------+---------+---------+ 1600
     GTTCCGCGCCAGCCCCTACGCGTACCCGGTACTGGCACACGACCTAGTGT

      K  A  R  S  G  M  R  M  G  H  D  R  V  L  D  H  M

     TGTTCCTCGACGGGTTGGAGGACGCCTATGACAAGGGCCGCCTGATGGGC
1601 ---------+---------+---------+---------+---------+ 1650
     ACAAGGAGCTGCCCAACCTCCTGCGGATACTGTTCCCGGCGGACTACCCG

       F  L  D  G  L  E  D  A  Y  D  K  G  R  L  M  G

     ACCTTCGCCGAGGATTGCGCCGGCGATCACGGTTTCACCCGCGAGGCGCA
1651 ---------+---------+---------+---------+---------+ 1700
     TGGAAGCGGCTCCTAACGCGGCCGCTAGTGCCAAAGTGGGCGCTCCGCGT

     T  F  A  E  D  C  A  G  D  H  G  F  T  R  E  A  Q

     GGACGACTATGCGCTGACCAGCCTGGCCCGCGCGCAGGACGCCATCGCCA
1701 ---------+---------+---------+---------+---------+ 1750
     CCTGCTGATACGCGACTGGTCGGACCGGGCGCGCGTCCTGCGGTAGCGGT

      D  D  Y  A  L  T  S  L  A  R  A  Q  D  A  I  A  S

     GCGGTGCCTTCGCCGCCGAGATCGCGCCCGTGACCGTCACGGCACGCAAG
1751 ---------+---------+---------+---------+---------+ 1800
     CGCCACGGAAGCGGCGGCTCTAGCGCGGGCACTGGCAGTGCCGTGCGTTC

       G  A  F  A  A  E  I  A  P  V  T  V  T  A  R  K
```

Fig. 7/4

```
     GTGCAGACCACCGTCGATACCGACGAGATGCCCGGCAAGGCCCGCCCCGA
1801 ---------+---------+---------+---------+---------+ 1850
     CACGTCTGGTGGCAGCTATGGCTGCTCTACGGGCCGTTCCGGGCGGGGCT

     V  Q  T  T  V  D  T  D  E  M  P  G  K  A  R  P  E

     GAAGATCCCCCATCTGAAGCCCGCCTTCCGTGACGGTGGCACGGTCACGG
1851 ---------+---------+---------+---------+---------+ 1900
     CTTCTAGGGGGTAGACTTCGGGCGGAAGGCACTGCCACCGTGCCAGTGCC

      K  I  P  H  L  K  P  A  F  R  D  G  G  T  V  T  A

     CGGCGAACAGCTCGTCGATCTCGGACGGGGCGGCGGCGCTGGTGATGATG
1901 ---------+---------+---------+---------+---------+ 1950
     GCCGCTTGTCGAGCAGCTAGAGCCTGCCCCGCCGCCGCGACCACTACTAC

       A  N  S  S  S  I  S  D  G  A  A  A  L  V  M  M

     CGCCAGTCGCAGGCCGAGAAGCTGGGCCTGACGCCGATCGCGCGGATCAT
1951 ---------+---------+---------+---------+---------+ 2000
     GCGGTCAGCGTCCGGCTCTTCGACCCGGACTGCGGCTAGCGCGCCTAGTA

     R  Q  S  Q  A  E  K  L  G  L  T  P  I  A  R  I  I

     CGGTCATGCGACCCATGCCGACCGTCCCGGCCTGTTCCCGACGGCCCCCA
2001 ---------+---------+---------+---------+---------+ 2050
     GCCAGTACGCTGGGTACGGCTGGCAGGGCCGGACAAGGGCTGCCGGGGGT

      G  H  A  T  H  A  D  R  P  G  L  F  P  T  A  P  I

     TCGGCGCGATGCGCAAGCTGCTGGACCGCACGGACACCCGCCTTGGCGAT
2051 ---------+---------+---------+---------+---------+ 2100
     AGCCGCGCTACGCGTTCGACGACCTGGCGTGCCTGTGGGCGGAACCGCTA

       G  A  M  R  K  L  L  D  R  T  D  T  R  L  G  D

     TACGACCTGTTCGAGGTGAACGAGGCATTCGCCGTCGTCGCCATGATCGC
2101 ---------+---------+---------+---------+---------+ 2150
     ATGCTGGACAAGCTCCACTTGCTCCGTAAGCGGCAGCAGCGGTACTAGCG

     Y  D  L  F  E  V  N  E  A  F  A  V  V  A  M  I  A

     GATGAAGGAGCTTGGCCTGCCACACGATGCCACGAACATCAACGGCGGGG
2151 ---------+---------+---------+---------+---------+ 2200
     CTACTTCCTCGAACCGGACGGTGTGCTACGGTGCTTGTAGTTGCCGCCCC

      M  K  E  L  G  L  P  H  D  A  T  N  I  N  G  G  A

     CCTGCGCGCTTGGGCATCCCATCGGCGCGTCGGGGGCGCGGATCATGGTC
2201 ---------+---------+---------+---------+---------+ 2250
     GGACGCGCGAACCCGTAGGGTAGCCGCGCAGCCCCCGCGCCTAGTACCAG

       C  A  L  G  H  P  I  G  A  S  G  A  R  I  M  V

     ACGCTGCTGAACGCGATGGCGGCGCGGGGCGCGACGCGCGGGGCCGCATC
2251 ---------+---------+---------+---------+---------+ 2300
     TGCGACGACTTGCGCTACCGCCGCGCCCCGCGCTGCGCGCCCCGGCGTAG

     T  L  L  N  A  M  A  A  R  G  A  T  R  G  A  A  S

     CGTCTGCATCGGCGGGGGCGAGGCGACGGCCATCGCGCTGGAACGGCTGA
2301 ---------+---------+---------+---------+---------+ 2350
     GCAGACGTAGCCGCCCCCGCTCCGCTGCCGGTAGCGCGACCTTGCCGACT

      V  C  I  G  G  G  E  A  T  A  I  A  L  E  R  L  S

     GCTAATTCATTTGCGCGAATCCGCGTTTTTCGTGCACGATGGGGGAACCG
2351 ---------+---------+---------+---------+---------+ 2400
     CGATTAAGTAAACGCGCTTAGGCGCAAAAAGCACGTGCTACCCCCTTGGC

       *
```

Fig. 7/5

```
                 .         .         .         .         .
2401   GAAACGGCCACGCCTGTTGTGGTTGCGTCGACCTGTCTTCGGGCCATGCC   2450
       ---------+---------+---------+---------+---------+
       CTTTGCCGGTGCGGACAACACCAACGCAGCTGGACAGAAGCCCGGTACGG

                 .         .         .         .         .
2451   CGTGACGCGATGTGGCAGGCGCATGGGGCGTTGCCGATCCGGTCGCATGA   2500
       ---------+---------+---------+---------+---------+
       GCACTGCGCTACACCGTCCGCGTACCCCGCAACGGCTAGGCCAGCGTACT

                 .         .         .         .         .
2501   CTGACGCAACGAAGGCACCGATGACGCCCAAGCAGCAATTCCCCCTACGC   2550
       ---------+---------+---------+---------+---------+
       GACTGCGTTGCTTCCGTGGCTACTGCGGGTTCGTCGTTAAGGGGGATGCG

             crtE -->  M  T  P  K  Q  Q  F  P  L  R

                 .         .         .         .         .
2551   GATCTGGTCGAGATCAGGCTGGCGCAGATCTCGGGCCAGTTCGGCGTGGT   2600
       ---------+---------+---------+---------+---------+
       CTAGACCAGCTCTAGTCCGACCGCGTCTAGAGCCCGGTCAAGCCGCACCA

       D  L  V  E  I  R  L  A  Q  I  S  G  Q  F  G  V  V

                 .         .         .         .         .
2601   CTCGGCCCCGCTCGGCGCGGCCATGAGCGATGCCGCCCTGTCCCCCGGCA   2650
       ---------+---------+---------+---------+---------+
       GAGCCGGGGCGAGCCGCGCCGGTACTCGCTACGGCGGGACAGGGGGCCGT

        S  A  P  L  G  A  A  M  S  D  A  A  L  S  P  G  K

                 .         .         .         .         .
2651   AACGCTTTCGCGCCGTGCTGATGCTGATGGTCGCCGAAAGCTCGGGCGGG   2700
       ---------+---------+---------+---------+---------+
       TTGCGAAAGCGCGGCACGACTACGACTACCAGCGGCTTTCGAGCCCGCCC

         R  F  R  A  V  L  M  L  M  V  A  E  S  S  G  G

                 .         .         .         .         .
2701   GTCTGCGATGCGATGGTCGATGCCGCCTGCGCGGTCGAGATGGTCCATGC   2750
       ---------+---------+---------+---------+---------+
       CAGACGCTACGCTACCAGCTACGGCGGACGCGCCAGCTCTACCAGGTACG

       V  C  D  A  M  V  D  A  A  C  A  V  E  M  V  H  A

                 .         .         .         .         .
2751   CGCATCGCTGATCTTCGACGACATGCCCTGCATGGACGATGCCAGGACCC   2800
       ---------+---------+---------+---------+---------+
       GCGTAGCGACTAGAAGCTGCTGTACGGGACGTACCTGCTACGGTCCTGGG

        A  S  L  I  F  D  D  M  P  C  M  D  D  A  R  T  R

                 .         .         .         .         .
2801   GTCGCGGTCAGCCCGCCACCCATGTCGCCCATGGCGAGGGGCGCGCGGTG   2850
       ---------+---------+---------+---------+---------+
       CAGCGCCAGTCGGGCGGTGGGTACAGCGGGTACCGCTCCCCGCGCGCCAC

         R  G  Q  P  A  T  H  V  A  H  G  E  G  R  A  V

                 .         .         .         .         .
2851   CTTGCGGGCATCGCCCTGATCACCGAGGCCATGCGGATTTTGGGCGAGGC   2900
       ---------+---------+---------+---------+---------+
       GAACGCCCGTAGCGGGACTAGTGGCTCCGGTACGCCTAAAACCCGCTCCG

       L  A  G  I  A  L  I  T  E  A  M  R  I  L  G  E  A

                 .         .         .         .         .
2901   GCGCGGCGCGACGCCGGATCAGCGCGCAAGGCTGGTCGCATCCATGTCGC   2950
       ---------+---------+---------+---------+---------+
       CGCGCCGCGCTGCGGCCTAGTCGCGCGTTCCGACCAGCGTAGGTACAGCG

        R  G  A  T  P  D  Q  R  A  R  L  V  A  S  M  S  R

                 .         .         .         .         .
2951   GCGCGATGGGACCGGTGGGGCTGTGCGCAGGGCAGGATCTGGACCTGCAC   3000
       ---------+---------+---------+---------+---------+
       CGCGCTACCCTGGCCACCCCGACACGCGTCCCGTCCTAGACCTGGACGTG

         A  N  G  P  V  G  L  C  A  G  Q  D  L  D  L  H
```

Fig. 7/6

```
      GCCCCCAAGGACGCCGCCGGGATCGAACGTGAACAGGACCTCAAGACCGG
3001  ---------+---------+---------+---------+---------+  3050
      CGGGGGTTCCTGCGGCGGCCCTAGCTTGCACTTGTCCTGGAGTTCTGGCC

      A P K D A A G I E R E Q D L K T G

      CGTGCTGTTCGTCGCGGGCCTCGAGATGCTGTCCATTATTAAGGGTCTGG
3051  ---------+---------+---------+---------+---------+  3100
      GCACGACAAGCAGCGCCCGGAGCTCTACGACAGGTAATAATTCCCAGACC

      V L F V A G L E M L S I I K G L D

      ACAAGGCCGAGACCGAGCAGCTCATGGCCTTCGGGCGTCAGCTTGGTCGG
3101  ---------+---------+---------+---------+---------+  3150
      TGTTCCGGCTCTGGCTCGTCGAGTACCGGAAGCCCGCAGTCGAACCAGCC

      K A E T E Q L M A F G R Q L G R

      GTCTTCCAGTCCTATGACGACCTGCTGGACGTGATCGGCGACAAGGCCAG
3151  ---------+---------+---------+---------+---------+  3200
      CAGAAGGTCAGGATACTGCTGGACGACCTGCACTAGCCGCTGTTCCGGTC

      V F Q S Y D D L L D V I G D K A S

      CACCGGCAAGGATACGGCGCGCGACACCGCCGCCCCCGGCCCAAAGGGCG
3201  ---------+---------+---------+---------+---------+  3250
      GTGGCCGTTCCTATGCCGCGCGCTGTGGCGGCGGGGGCCGGGTTTCCCGC

      T G K D T A R D T A A P G P K G G

      GCCTGATGGCGGTCGGACAGATGGGCGACGTGGCGCAGCATTACCGCGCC
3251  ---------+---------+---------+---------+---------+  3300
      CGGACTACCGCCAGCCTGTCTACCCGCTGCACCGCGTCGTAATGGCGCGG

      L M A V G Q M G D V A Q H Y R A

      AGCCGCGCGCAACTGGACGAGCTGATGCGCACCCGGCTGTTCCGCGGGGG
3301  ---------+---------+---------+---------+---------+  3350
      TCGGCGCGCGTTGACCTGCTCGACTACGCGTGGGCCGACAAGGCGCCCCC

      S R A Q L D E L M R T R L F R G G

      GCAGATCGCGGACCTGCTGGCCCGCGTGCTGCCGCATGACATCCGCCGCA
3351  ---------+---------+---------+---------+---------+  3400
      CGTCTAGCGCCTGGACGACCGGGCGCACGACGGCGTACTGTAGGCGGCGT

      Q I A D L L A R V L P H D I R R S

      GCGCCTAGGCGCGCGGTCGGGTCCACAGGCCGTCGCGGCTGATTTCGCCG
3401  ---------+---------+---------+---------+---------+  3450
      CGCGGATCCGCGCGCCAGCCCAGGTGTCCGGCAGCGCCGACTAAAGCGGC

      A *
        * A R P R T W L G D R S I E G

      CCGCGCAGGCGCGATGCGGCCGCGTCCAAGCCTCCGCGCGCCAGAAGCCC
3451  ---------+---------+---------+---------+---------+  3500
      GGCGCGTCCGCGCTACGCCGGCGCAGGTTCGGAGGCGCGCGGTCTTCGGG

      G R L R S A A A D L G G R A L L G

      GATCTTGGCAGCCTTCGACGTGCTGATCCGCTGGCGATAGGCCTCGGGGC
3501  ---------+---------+---------+---------+---------+  3550
      CTAGAACCGTCGGAAGCTGCACGACTAGGCGACCGCTATCCGGAGCCCCG

      I K A A K S T S I R Q R Y A E P

      CACCCTGCCGGATGCGCGTCCCGATTGCGCGATAGATACGCAGCGCGGCG
3551  ---------+---------+---------+---------+---------+  3600
      GTGGGACGGCCTACGCGCAGGGCTAACGCGCTATCTATGCGTCGCGCCGC
      G G Q R I R T G I A R Y I R L A A
```

Fig. 7/7

```
     GCGATCGACCACGCGCAGCGCGGCGGCAGATGCGGAAGCCCCTGCCGCGC
3601 ---------+---------+---------+---------+---------+ 3650
     CGCTAGCTGGTGCGCGTCGCGCCGCCGTCTACGCCTTCGGGGACGGCGCG

      A  I  S  W  A  C  R  P  P  L  H  P  L  G  Q  R  A

     CGAGGCATAATAGGGCTCGGCCGCGTCAAGCAGGCGGATGATGACGGAAT
3651 ---------+---------+---------+---------+---------+ 3700
     GCTCCGTATTATCCCGAGCCGGCGCAGTTCGTCCGCCTACTACTGCCTTA

       S  A  Y  Y  P  E  A  A  D  L  L  R  I  I  V  S

     AGAGCGCGTCCGAAGGCACCGGACCCTCAACCGTCGCCCCCGCCTCGGCC
3701 ---------+---------+---------+---------+---------+ 3750
     TCTCGCGCAGGCTTCCGTGGCCTGGGAGTTGGCAGCGGGGGCGGAGCCGG

     Y  L  A  D  S  P  V  P  G  E  V  T  A  G  A  E  A

     AGCCAGTCGGCAGGCAGATAGCAGCGCCCGATGGCGGCATCGTCGATCAC
3751 ---------+---------+---------+---------+---------+ 3800
     TCGGTCAGCCGTCCGTCTATCGTCGCGGGCTACCGCCGTAGCAGCTAGTG

      L  W  D  A  P  L  Y  C  R  G  I  A  A  D  D  I  V

     GTCGCGAGCGATGTTCGTCAGCTGGAACGCAAGGCCCAGATCGCAGGCGC
3801 ---------+---------+---------+---------+---------+ 3850
     CAGCGCTCGCTACAAGCAGTCGACCTTGCGTTCCGGGTCTAGCGTCCGCG

       D  R  A  I  N  T  L  Q  F  A  L  G  L  D  C  A

     GATCCAGCACCGCATCGTCCTGCACGCCCATCACCCGCGCCATCATCACG
3851 ---------+---------+---------+---------+---------+ 3900
     CTAGGTCGTGGCGTAGCAGGACGTGCGGGTAGTGGGCGCGGTAGTAGTGC
     R  D  L  V  A  D  D  Q  V  G  M  V  R  A  M  M  V

     CCCACGACCCCCGCGACGTGGTAGGAATATTCCAGCACGTCATCCAGGCT
3901 ---------+---------+---------+---------+---------+ 3950
     GGGTGCTGGGGGCGCTGCACCATCCTTATAAGGTCGTGCAGTAGGTCCGA

      G  V  V  G  A  V  R  Y  S  Y  E  L  V  D  D  L  S

     GCGGTATTCGCGATCCGCGACATCCATCGCGAAACCCTCGATCAGGTCCA
3951 ---------+---------+---------+---------+---------+ 4000
     CGCCATAAGCGCTAGGCGCTGTAGGTAGCGCTTTGGGAGCTAGTCCAGGT

       R  Y  E  R  D  A  V  D  M  A  F  G  E  I  L  D

     TCGGCCAAAGGTCCGGGAAATCATGCCGCCGGGCGACCTGGCGCAGCGCC
4001 ---------+---------+---------+---------+---------+ 4050
     AGCCGGTTTCCAGGCCCTTTAGTACGGCGGCCCGCTGGACCGCGTCGCGG

     M  P  W  L  D  P  F  D  H  R  R  A  V  Q  R  L  A

     GCGAAGGGCGGCGACATCGGGCCGTCCTCGTGCAGCGCGGCCAGCGTGTC
4051 ---------+---------+---------+---------+---------+ 4100
     CGCTTCCCGCCGCTGTAGCCCGGCAGGAGCACGTCGCGCCGGTCGCACAG

      A  F  P  P  S  M  P  G  D  E  H  L  A  A  L  T  D

     GGCGCGCAGCGCCCCCAGCCGCGCCTGTGGGTCGCCGCCCGCCTCGGGGG
4101 ---------+---------+---------+---------+---------+ 4150
     CCGCGCGTCGCGGGGGTCGGCGCGGACACCCAGCGGCGGGCGGAGCCCCC

       A  R  L  A  G  L  R  A  Q  P  D  G  G  A  E  P

     CAGAACCCATCACCTGCCCGTCGATCACGTCATCCGCATGCCTGCACCAG
4151 ---------+---------+---------+---------+---------+ 4200
     GTCTTGGGTAGTGGACGGGCAGCTAGTGCAGTAGGCGTACGGACGTGGTC
     A  S  G  M  V  Q  G  D  I  V  D  D  A  H  R  C  W
```

Fig. 7/8

```
4201 GCATAGAGCATGACCGTATCCTCGCGGATGCCGGGCGGCATCAGCTTGGC 4250
     ---------+---------+---------+---------+---------+
     CGTATCTCGTACTGGCATAGGAGCGCCTACGGCCCGCCGTAGTCGAACCG

      A  Y  L  N  V  T  D  E  R  I  G  P  P  M  L  K  A

4251 CGCCTGCGCGAAGCTTTGCGAACCCTGCGCGATGGCCGCTTCGGAAGTCG 4300
     ---------+---------+---------+---------+---------+
     GCGGACGCGCTTCGAAACGCTTGGGACGCGCTACCGGCGAAGCCTTCAGC

       A  Q  A  F  S  Q  S  G  Q  A  I  A  A  E  S  T

4301 CCGTCAGATCGGTCATGCGACGGCCAGGTCCGACAGCATGACCTGCGCCG 4350
     ---------+---------+---------+---------+---------+
     GGCAGTCTAGCCAGTACGCTGCCGGTCCAGGCTGTCGTACTGGACGCGGC

                 *  A  V  A  L  D  S  L  M  V  Q  A
     A  T  L  D  T  M
         <--  crtB

4351 TGGCCTTGGCGCTGCCAACGACACCCGGGATGCCCGCACCCGGATGCGTG 4400
     ---------+---------+---------+---------+---------+
     ACCGGAACCGCGACGGTTGCTGTGGGCCCTACGGGCGTGGGCCTACGCAC

     T  A  K  A  S  G  V  V  G  P  I  G  A  G  P  H  T

4401 CCCGCCCCCACGATGTAGAAGTTCGGGATCGCGCGGTCGCGGTTATGCGG 4450
     ---------+---------+---------+---------+---------+
     GGGCGGGGGTGCTACATCTTCAAGCCCTAGCGCGCCAGCGCCAATACGCC

      G  A  G  V  I  Y  F  N  P  I  A  R  D  R  N  H  P

4451 GCGGAACCAGGCGGATTGCGTCAGGATCGGCTCGACCGAGAAGGCGCTGC 4500
     ---------+---------+---------+---------+---------+
     CGCCTTGGTCCGCCTAACGCAGTCCTAGCCGAGCTGGCTCTTCCGCGACG

       R  F  W  A  S  Q  T  L  I  P  E  V  S  F  A  S

4501 CGTGATGGGCCGACAGTTCGGTGCTGAAATCGGCGGGGCTGAAGATGCGG 4550
     ---------+---------+---------+---------+---------+
     GCACTACCCGGCTGTCAAGCCACGACTTTAGCCGCCCCGACTTCTACGCC

     G  H  H  A  S  L  E  T  S  F  D  A  P  S  F  I  R

4551 CTGACGGTCAGGTGCTTGCGCAGGTCGGGGATGGCGCGGCGCTCCAGTTC 4600
     ---------+---------+---------+---------+---------+
     GACTGCCAGTCCACGAACGCGTCCAGCCCCTACCGCGCCGCGAGGTCAAG

      S  V  T  L  H  K  R  L  D  P  I  A  R  R  E  L  E

4601 CTCGAAGATGCGCTCGGCATAGCCCGGGGCCTCGGCTTCCCAATCGACAT 4650
     ---------+---------+---------+---------+---------+
     GAGCTTCTACGCGAGCCGTATCGGGCCCCGGAGCCGAAGGGTTAGCTGTA

       E  F  I  R  E  A  Y  G  P  A  E  A  E  W  D  V

4651 CGGCGCGGCCCAGATGCGGAACGGGCGCAAGGACGTAATGCGTGGACATC 4700
     ---------+---------+---------+---------+---------+
     GCCGCGCCGGGTCTACGCCTTGCCCGCGTTCCTGCATTACGCACCTGTAG

     D  A  R  G  L  H  P  V  P  A  L  V  Y  H  T  S  M

4701 CCCTCGGGGGCCAGGCTGGGATCGGTCACGCAGGGCGAATGCAGATACAT 4750
     ---------+---------+---------+---------+---------+
     GGGAGCCCCCGGTCCGACCCTAGCCAGTGCGTCCCGCTTACGTCTATGTA

      G  E  P  A  L  S  P  D  T  V  C  P  S  H  L  Y  M

4751 CGAGAAATCGTCCGGCAGGCGTGGCCCGTTGAAGATCTCGTTCACCAGCC 4800
     ---------+---------+---------+---------+---------+
     GCTCTTTAGCAGGCCGTCCGCACCGGGCAACTTCTAGAGCAAGTGGTCGG

       S  F  D  D  P  L  R  P  G  N  F  I  E  N  V  L
```

Fig. 7/9

```
      CCTTGTAGCGCGGGCCGAAGATGACGCTGTGGTGGGCCAGGTTCTCGGGG
4801  ---------+---------+---------+---------+---------+  4850
      GGAACATCGCGCCCGGCTTCTACTGCGACACCACCCGGTCCAAGAGCCCC

      G  K  Y  R  P  G  F  I  V  S  H  H  A  L  N  E  P

      CGCTTGGACAGGCCGAAATGCAGCACGAACAGCGACATCGACCAGCGCTG
4851  ---------+---------+---------+---------+---------+  4900
      GCGAACCTGTCCGGCTTTACGTCGTGCTTGTCGCTGTAGCTGGTCGCGAC

       R  K  S  L  G  F  H  L  V  F  L  S  M  S  W  R  Q

      CCGGTTCAGGATCGCGGCCTTGGTGCGCCCGCGGCGGGTATGGCCCAGCA
4901  ---------+---------+---------+---------+---------+  4950
      GGCCAAGTCCTAGCGCCGGAACCACGCGGGCGCCGCCCATACCGGGTCGT

        R  N  L  I  A  A  K  T  R  G  R  R  T  H  G  L

      GGTCGCGATAGCTGTGCATCACGTCGCCGTTGCTGGCCACCGTATCCGCG
4951  ---------+---------+---------+---------+---------+  5000
      CCAGCGCTATCGACACGTAGTGCAGCGGCAACGACCGGTGGCATAGGCGC

      L  D  R  Y  S  H  M  V  D  G  N  S  A  V  T  D  A

      CGCAACTGCCGCCCGTCCAGCAGCGTGACGCCCGTGGCGCGATCGCCCTC
5001  ---------+---------+---------+---------+---------+  5050
      GCGTTGACGGCGGGCAGGTCGTCGCACTGCGGGCACCGCGCTAGCGGGAG

       R  L  Q  R  G  D  L  L  T  V  G  T  A  R  D  G  E

      GGTGTCGATCCGCGTGACGCGGGCATTCAGCAGCAGCGTGCCGCCAAGAC
5051  ---------+---------+---------+---------+---------+  5100
      CCACAGCTAGGCGCACTGCGCCCGTAAGTCGTCGTCGCACGGCGGTTCTG

        T  D  I  R  T  V  R  A  N  L  L  L  T  G  G  L

      GCTCGAACAGGGCGACCATGCCCGCGACCAGCTGGTTGGTGCCGCCCTTG
5101  ---------+---------+---------+---------+---------+  5150
      CGAGCTTGTCCCGCTGGTACGGGCGCTGGTCGACCAACCACGGCGGGAAC

      R  E  F  L  A  V  M  G  A  V  L  Q  N  T  G  G  K

      GCGAACCAGACGCCGCCGCGCCGTTCCAGCGCATGGATCAGCGCATAGAT
5151  ---------+---------+---------+---------+---------+  5200
      CGCTTGGTCTGCGGCGGCGCGGCAAGGTCGCGTACCTAGTCGCGTATCTA

       A  F  W  V  G  G  R  R  E  L  A  H  I  L  A  Y  I

      CGAGCTGGTCGAAAACGGGTTCCCGCCGACCAGCAGCGTGTGGAACGAGA
5201  ---------+---------+---------+---------+---------+  5250
      GCTCGACCAGCTTTTGCCCAAGGGCGGCTGGTCGTCGCACACCTTGCTCT

        S  S  T  S  F  P  N  G  G  V  L  L  T  H  F  S

      AGGCCTGCCGCAGATGCGGGTCCTGGATGAAGCGCGCCACCATGCTGTGG
5251  ---------+---------+---------+---------+---------+  5300
      TCCGGACGGCGTCTACGCCCAGGACCTACTTCGCGCGGTGGTACGACACC

      F  A  Q  R  L  H  P  D  Q  I  F  R  A  V  M  S  H

      ACCGAGCGGTATGCCTGCAGGCGCATCAGCGCCGGCGCGGCGTTCAGCAT
5301  ---------+---------+---------+---------+---------+  5350
      TGGCTCGCCATACGGACGTCCGCGTAGTCGCGGCCGCGCCGCAAGTCGTA

       V  S  R  Y  A  Q  L  R  M  L  A  P  A  A  N  L  M

      CTGGCCCAGCTTCAGGAAGGGCGTGGTCCCCAGCTTCAGATACCCCTCGC
5351  ---------+---------+---------+---------+---------+  5400
      GACCGGGTCGAAGTCCTTCCCGCACCAGGGGTCGAAGTCTATGGGGAGCG

        Q  G  L  K  L  F  P  T  T  G  L  K  L  Y  G  E
```

Fig. 7/10

```
5401 GATAGACCTCCTCGGCGTAATCGTGGAAGCGGCGATAGCCATCGACATCG 5450
     ---------+---------+---------+---------+---------+
     CTATCTGGAGGAGCCGCATTAGCACCTTCGCCGCTATCGGTAGCTGTAGC

     R  Y  V  E  E  A  Y  D  H  F  R  R  Y  G  D  V  D

5451 GCGGGATTGAAGGAGGCGACCTGGCGGATCAGCTCGTCGTCGTCGTTCAC 5500
     ---------+---------+---------+---------+---------+
     CGCCCTAACTTCCTCCGCTGGACCGCCTAGTCGAGCAGCAGCAGCAAGTG

      A  P  N  F  S  A  V  Q  R  I  L  E  D  D  D  N  V

5501 GTATTCGAAGCTGCGGCCGTCCGCCCATGTCAGCCGGTAGAAGGGCGAGA 5550
     ---------+---------+---------+---------+---------+
     CATAAGCTTCGACGCCGGCAGGCGGGTACAGTCGGCCATCTTCCCGCTCT

       Y  E  F  S  R  G  D  A  W  T  L  R  Y  F  P  S

5551 CCGGCAGCAGCGTCACGTCACGCTCCATCGGTTGGCCGCTGAGGGCCCAC 5600
     ---------+---------+---------+---------+---------+
     GGCCGTCGTCGCAGTGCAGTGCGAGGTAGCCAACCGGCGACTCCCGGGTG

     V  P  L  L  T  V  D  R  E  M  P  Q  G  S  L  A  W

5601 AGCTCTCGCAGGCTGTCGGGGTCGGTCACGACCGTCGGGCCTGCATCGAA 5650
     ---------+---------+---------+---------+---------+
     TCGAGAGCGTCCGACAGCCCCAGCCAGTGCTGGCAGCCCGGACGTAGCTT

      L  E  R  L  S  D  P  D  T  V  V  T  P  G  A  D  F

5651 GACGTGGCCCTGATCGTTCCAGACATAGGCGCGGCCGCCGGGCTTGTCGC 5700
     ---------+---------+---------+---------+---------+
     CTGCACCGGGACTAGCAAGGTCTGTATCCGCGCCGGCGGCCCGAACAGCG

       V  H  G  Q  D  N  W  V  Y  A  R  G  G  P  K  D

5701 GGGCCTCGACGATGGTGGTCGCGATGCCGGCCGATTGCAGGCGGATGGCA 5750
     ---------+---------+---------+---------+---------+
     CCCGGAGCTGCTACCACCAGCGCTACGGCCGGCTAACGTCCGCCTACCGT

     R  A  E  V  I  T  T  A  I  G  A  S  Q  L  R  I  A

5751 AGCGCAAGCCCGCCGAAACCTGCGCCGATGACGATGGCGGAACTCATGCT 5800
     ---------+---------+---------+---------+---------+
     TCGCGTTCGGGCGGCTTTGGACGCGGCTACTGCTACCGCCTTGAGTACGA

      L  A  L  G  G  F  G  A  G  I  V  I  A  S  S  M  <-- crtI
                                                   *  A

5801 CTCTCCTGCAGCAGGGGGCGTTCGGGCAGGCAGCGCACGGCCTGCGACAG 5850
     ---------+---------+---------+---------+---------+
     GAGAGGACGTCGTCCCCCGCAAGCCCGTCCGTCGCGTGCCGGACGCTGTC

      R  E  Q  L  L  P  R  E  P  L  C  R  V  A  Q  S  L

5851 CGGAATGGGCGGGCGTCCGGTGACGATGCGAAGCCGGTCGGCCAATGTCA 5900
     ---------+---------+---------+---------+---------+
     GCCTTACCCGCCCGCAGGCCACTGCTACGCTTCGGCCAGCCGGTTACAGT

       P  I  P  P  R  G  T  V  I  R  L  R  D  A  L  T

5901 GGCGCCCGGCATAGAAGCGCTCGATCAGCGGCTGCGGCAGGCGGTAGAAC 5950
     ---------+---------+---------+---------+---------+
     CCGCGGGCCGTATCTTCGCGAGCTAGTCGCCGACGCCGTCCGCCATCTTG

     L  R  G  A  Y  F  R  E  I  L  P  Q  P  L  R  Y  F

5951 CGCTGCAGCAGGCGATAGCGACGGTCGGGCGGGCAGCCGCGGAACAGCAT 6000
     ---------+---------+---------+---------+---------+
     GCGACGTCGTCCGCTATCGCTGCCAGCCCGCCCGTCGGCGCCTTGTCGTA
      R  Q  L  L  R  Y  R  R  D  P  P  C  G  R  F  L  M
```

Fig. 7/11

```
6001 CCGGTTCAGCAGCCGCAGGAAGCGGTCGCGATCCGCGCGATCGATGGCCC 6050
     ---------+---------+---------+---------+---------+
     GGCCAAGTCGTCGGCGTCCTTCGCCAGCGCTAGGCGCGCTAGCTACCGGG

      R  N  L  L  P  L  F  R  D  R  D  A  R  D  I  A

6051 AGCCGCGCACCGCGCGACGGGCGGACGCGGTCGTCAGGTCGCGCGCCGCG 6100
     ---------+---------+---------+---------+---------+
     TCGGCGCGTGGCGCGCTGCCCGCCTGCGCCAGCAGTCCAGCGCGCGGCGC

     W  G  R  V  A  R  R  A  S  A  T  T  L  D  R  A  A

6101 ATGGCATCCGCGACCTGCGCGGCATAGGGCAGCGAATATCCGGTGACGGG 6150
     ---------+---------+---------+---------+---------+
     TACCGTAGGCGCTGGACGCGCCGTATCCCGTCGCTTATAGGCCACTGCCC

      I  A  D  A  V  Q  A  A  Y  P  L  S  Y  G  T  V  P

6151 GTGGAACAGCCCTGCCCCCAGCCCAACCGGCACCGCCCCCTGCGCGTGGT 6200
     ---------+---------+---------+---------+---------+
     CACCTTGTCGGGACGGGGGTCGGGTTGGCCGTGGCGGGGGACGCGCACCA

      H  F  L  G  A  G  L  G  V  P  V  A  G  Q  A  H

6201 CGCGCCAGAAGCCTATGGCGTCATGGGCCAGCGCGATGGGCAGGATGCCC 6250
     ---------+---------+---------+---------+---------+
     GCGCGGTCTTCGGATACCGCAGTACCCGGTCGCGCTACCCGTCCTACGGG

     D  R  W  F  G  I  A  D  H  A  L  A  I  P  L  I  G

6251 CTTTCGCGCCGCATCTCCTGCCCGGTCCAGCCCCGCCTGGCGGCATAGTC 6300
     ---------+---------+---------+---------+---------+
     GAAAGCGCGGCGTAGAGGACGGGCCAGGTCGGGGCGGACCGCCGTATCAG
     R  E  R  R  M  E  Q  G  T  W  G  R  R  A  A  Y  D

6301 CAGCGACGCCTGCGCCAGCGCGCCATCGTCCAGATCGCCGCCGTCGCTGT 6350
     ---------+---------+---------+---------+---------+
     GTCGCTGCGGACGCGGTCGCGCGGTAGCAGGTCTAGCGGCGGCAGCGACA

      L  S  A  Q  A  L  A  G  D  D  L  D  G  G  D  S

6351 AGCGCGTATCCTCGATCAGGATGCGGGTGGGACTGAAGGGCAGCAGATAG 6400
     ---------+---------+---------+---------+---------+
     TCGCGCATAGGAGCTAGTCCTACGCCCACCCTGACTTCCCGTCGTCTATC

     Y  R  T  D  E  I  L  I  R  T  P  S  F  P  L  L  Y

6401 ATGAAGCGGTACCCGTCCATCTGCGGAACGGTCGCGTCCATGATCATCGG 6450
     ---------+---------+---------+---------+---------+
     TACTTCGCCATGGGCAGGTAGACGCCTTGCCAGCGCAGGTACTAGTAGCC

      I  F  R  Y  G  D  M  Q  P  V  T  A  D  M  I  M  P

6451 GCGCTCGACGCCATGGGGGGCGTCGGTCTCGATCTCGACGCCCACGAATT 6500
     ---------+---------+---------+---------+---------+
     CGCGAGCTGCGGTACCCCCCGCAGCCAGAGCTAGAGCTGCGGGTGCTTAA

      R  E  V  G  H  P  A  D  T  E  I  E  V  G  V  F

6501 TCTGGAAACCCACGGTCAGGTGCGGGGTCTCGACGGCACCACGGGCGTCG 6550
     ---------+---------+---------+---------+---------+
     AGACCTTTGGGTGCCAGTCCACGCCCCAGAGCTGCCGTGGTGCCCGCAGC

     K  Q  F  G  V  T  L  H  P  T  E  V  A  G  R  A  D

6551 ATCACGCAGGCAGCCTCGATCCGCGAGCCGTCCGTCAGCGTCGCGCCGGT 6600
     ---------+---------+---------+---------+---------+
     TAGTGCGTCCGTCGGAGCTAGGCGCTCGGCAGGCAGTCGCAGCGCGGCCA
     I  V  C  A  A  E  I  R  S  G  D  T  L  T  A  G  T
```

Fig. 7/12

```
6601 ATCGTCCAGCGTCGCGACATGCGTATTCCACCGCAGATCGACACCCTGCA 6650
     ---------+---------+---------+---------+---------+
     TAGCAGGTCGCAGCGCTGTACGCATAAGGTGGCGTCTAGCTGTGGGACGT

      D  D  L  T  A  V  H  T  N  W  R  L  D  V  G  Q

6651 GCAGCCCGATCAGCGCGCCCGCCTCGATCGAGCCATAGCCTGTCGTCAGG 6700
     ---------+---------+---------+---------+---------+
     CGTCGGGCTAGTCGCGCGGGCGGAGCTAGCTCGGTATCGGACAGCAGTCC

     L  L  G  I  L  A  G  A  E  I  S  G  Y  G  T  T  L

6701 CGGCGCGAATGGTCGGGAAACGCGACCTCCTGATCCGTCCATTCGCCGCG 6750
     ---------+---------+---------+---------+---------+
     GCCGCGCTTACCAGCCCTTTGCGCTGGAGGACTAGGCAGGTAAGCGGCGC

      R  R  S  H  D  P  F  A  V  E  Q  D  T  W  E  G  R

6751 ACGAATGGGCGACAGGCGCGCCAGCCATTCGGGCGAAAGATCCGTGTCGT 6800
     ---------+---------+---------+---------+---------+
     TGCTTACCCGCTGTCCGCGCGGTCGGTAAGCCCGCTTTCTAGGCACAGCA

      R  I  P  S  L  R  A  L  W  E  P  S  L  D  T  D

6801 GGCAGGACCAGGTGTGCTGGTCCGAGGGGCCGGACCGCGCGTCGAGCATC 6850
     ---------+---------+---------+---------+---------+
     CCGTCCTGGTCCACACGACCAGGCTCCCCGGCCTGGCGCGCAGCTCGTAG

     H  C  S  W  T  H  Q  D  S  P  G  S  R  A  D  L  M

6851 ACGATGCGCGCATCCGGTCTGCGGTCGCGAACGGCAAGCGCGATCAGCGC 6900
     ---------+---------+---------+---------+---------+
     TGCTACGCGCGTAGGCCAGACGCCAGCGCTTGCCGTTCGCGCTAGTCGCG
      V  I  R  A  D  P  R  R  D  R  V  A  L  A  I  L  A

6901 ACCGGACAGCCCCGCGCCCGCGATCAGCAGATCATGGCTCATGTATTGCG 6950
     ---------+---------+---------+---------+---------+
     TGGCCTGTCGGGGCGCGGGCGCTAGTCGTCTAGTACCGAGTACATAACGC

                                               *  T  N  R
       G  S  L  G  A  G  A  I  L  L  D  H  S  M
                                      <-- crtY

6951 ATCCGCCCCTTCGCGGTCCTTCAGCAGCGCGCCCGAGCGTTTCAGCTCTG 7000
     ---------+---------+---------+---------+---------+
     TAGGCGGGGAAGCGCCAGGAAGTCGTCGCGCGGGCTCGCAAAGTCGAGAC

       D  A  G  E  R  D  K  L  L  A  G  S  R  K  L  E

7001 CCTTGAGGCTGTCGACCGAGGGCGCCCAGATGAAACCGAAGCTGACGCAG 7050
     ---------+---------+---------+---------+---------+
     GGAACTCCGACAGCTGGCTCCCGCGGGTCTACTTTGGCTTCGACTGCGTC

     A  K  L  S  D  V  S  P  A  W  I  F  G  F  S  V  C

7051 TTCTCGCGGCCATGGACCGCGTGATGCATCCTGTGTGCCTGGTAGACGCG 7100
     ---------+---------+---------+---------+---------+
     AAGAGCGCCGGTACCTGGCGCACTACGTAGGACACACGGACCATCTGCGC

       N  E  R  G  H  V  A  H  H  M  R  H  A  Q  Y  V  R

7101 ACGAAGATAGCCGCGCTTGGGGACATAGCGGAACGGCCAGCGCCCATGCA 7150
     ---------+---------+---------+---------+---------+
     TGCTTCTATCGGCGCGAACCCCTGTATCGCCTTGCCGGTCGCGGGTACGT

       R  L  Y  G  R  K  P  V  Y  R  F  P  W  R  G  H

7151 CCAAGCCGTCATGCAGGAAATAGTAGATCAGCCCGTAGCAGGTGACCCCC 7200
     ---------+---------+---------+---------+---------+
     GGTTCGGCAGTACGTCCTTTATCATCTAGTCGGGCATCGTCCACTGGGGG
     V  L  G  D  H  L  F  Y  Y  I  L  G  Y  C  T  V  G
```

Fig. 7/13

```
      ACCGCCAGCCACCAGGCCAGATCCGACCCCATCGCGCCGATCGCGAACAG
7201  ---------+---------+---------+---------+---------+  7250
      TGGCGGTCGGTGGTCCGGTCTAGGCTGGGGTAGCGCGGCTAGCGCTTGTC

       V  A  L  W  W  A  L  D  S  G  M  A  G  I  A  F  L

      CACGATCGAGATTACCGCGAAGATGACGCCATAGAGGTCGTTCTTCTCGA
7251  ---------+---------+---------+---------+---------+  7300
      GTGCTAGCTCTAATGGCGCTTCTACTGCGGTATCTCCAGCAAGAAGAGCT

        V  I  S  I  V  A  F  I  V  G  Y  L  D  N  K  E

      GCGCGTGGTCGTGATCCTCGTCGTGGTGCGATTTATGCCAGCCCCAGCCC
7301  ---------+---------+---------+---------+---------+  7350
      CGCGCACCAGCACTAGGAGCAGCACCACGCTAAATACGGTCGGGGTCGGG

      L  A  H  D  H  D  E  D  H  H  S  K  H  W  G  W  G

      AGGGGGCCATGCATGATCCACCGATGGACGGAGTAGGCCGTCAGCTCCAT
7351  ---------+---------+---------+---------+---------+  7400
      TCCCCCGGTACGTACTAGGTGGCTACCTGCCTCATCCGGCAGTCGAGGTA

       L  P  G  H  M  I  W  R  H  V  S  Y  A  T  L  E  M

      CGCGGCGACGGTCAGGATGACGGTCAGGATTGCGGCCCAAGTGCTCATGC
7401  ---------+---------+---------+---------+---------+  7450
      GCGCCGCTGCCAGTCCTACTGCCAGTCCTAACGCCGGGTTCACGAGTACG

        A  A  V  T  L  I  V  T  L  I  A  A  W  T  S  M
                                        <-- crtZ

      CGGCCCCTTGCTTGATATGACAGGGAACAGGCTACGCTGCCGCGCGGTGC
7451  ---------+---------+---------+---------+---------+  7500
      GCCGGGGAACGAACTATACTGTCCCTTGTCCGATGCGACGGCGCGCCACG

      ATGACCAGCCCATCGGGGTGCGACCAAAGGGCATCGCGTGACATCTGCGT
7501  ---------+---------+---------+---------+---------+  7550
      TACTGGTCGGGTAGCCCCACGCTGGTTTCCCGTAGCGCACTGTAGACGCA

      TCAGGGCTCATAGGCGGATCATCCGTGACATTCGCCGCCGAACGCGGCAG
7551  ---------+---------+---------+---------+---------+  7600
      AGTCCCGAGTATCCGCCTAGTAGGCACTGTAAGCGGCGGCTTGCGCCGTC

      GCGCATCACGCGTTCCGTCGCTGGAAATATTAATGTTTTCCCGAAGATGG
7601  ---------+---------+---------+---------+---------+  7650
      CGCGTAGTGCGCAAGGCAGCGACCTTTATAATTACAAAAGGGCTTCTACC

      TCGGGGCGAGAGGATTCGAACCTCCGACCTACGGTACCCAAAACCGTCGC
7651  ---------+---------+---------+---------+---------+  7700
      AGCCCCGCTCTCCTAAGCTTGGAGGCTGGATGCCATGGGTTTTGGCAGCG

      GCTACCAGGCTGCGCTACGCCCCGACTGCGGAAGGCTTTAGCCGATTGTT
7701  ---------+---------+---------+---------+---------+  7750
      CGATGGTCCGACGCGATGCGGGGCTGACGCCTTCCGAAATCGGCTAACAA

      CCGGCAAGGGAAAGACCTAGTCGCAGGCCAGGACCGCATTGTCGCCCATG
7751  ---------+---------+---------+---------+---------+  7800
      GGCCGTTCCCTTTCTGGATCAGCGTCCGGTCCTGGCGTAACAGCGGGTAC
      *  D  C  A  L  V  A  N  D  G  M
```

Fig. 7/14

```
7801 CCCGGATGCGCCATCGGCTGACCGGGCTTCAGGCCAAGGCGATCCGCCTC 7850
     ---------+---------+---------+---------+---------+
     GGGCCTACGCGGTAGCCGACTGGCCCGAAGTCCGGTTCCGCTAGGCGGAG

      G  P  H  A  N  P  Q  G  P  K  L  G  L  R  D  A  E

7851 TCCGCCCGCGATTTCGAGGACGAACAGCCGGTCGGGGTCCGGATCGCCGA 7900
     ---------+---------+---------+---------+---------+
     AGGCGGGCGCTAAAGCTCCTGCTTGTCGGCCAGCCCCAGGCCTAGCGGCT

      G  G  A  I  E  L  V  F  L  R  D  P  D  P  D  G

7901 CCGCCGCGCCCGGAATGGGCGTCTCGTCCAGCGGGCGCGCATTGCGGTGG 7950
     ---------+---------+---------+---------+---------+
     GGCGGCGCGGGCCTTACCCGCAGAGCAGGTCGCCCGCGCGTAACGCCACC

     V  A  A  G  P  I  P  T  E  D  L  P  R  A  N  R  H

7951 ATGTGGCGGATGACGCCGGTTTCATCCGCAAAGACCATGTCCAGCGGGAT 8000
     ---------+---------+---------+---------+---------+
     TACACCGCCTACTGCGGCCAAAGTAGGCGTTTCTGGTACAGGTCGCCCTA

      I  H  R  I  V  G  T  E  D  A  F  V  M  D  L  P  I

8001 CAGTGTGTTGCGCATCCAGAAGGACACCGGCTGGGGCGATTCGTAGATGA 8050
     ---------+---------+---------+---------+---------+
     GTCACACAACGCGTAGGTCTTCCTGTGGCCGACCCCGCTAAGCATCTACT

      L  T  N  R  M  W  F  S  V  P  Q  P  S  E  Y  I

8051 ACAGCATTCCGGTGCCCGCAGGCAGCTCCTTGCGGAACATCAGGCCCTGC 8100
     ---------+---------+---------+---------+---------+
     TGTCGTAAGGCCACGGGCGTCCGTCGAGGAACGCCTTGTAGTCCGGGACG

8101 GCGCGCTCTTCGGGGCTGTCCGCGACCTCGACCCGAAACCCGAGCGTTTC 8150
     ---------+---------+---------+---------+---------+
     CGCGCGAGAAGCCCCGACAGGCGCTGGAGCTGGGCTTTGGGCTCGCAAAG

      A  R  E  E  P  S  D  A  V  E  V  R  F  G  L  T  E

8151 CGCACCGGTATCGACGACAAGACTGCCGGGCGCGCATTCCACCGCCGCCG 8200
     ---------+---------+---------+---------+---------+
     GCGTGGCCATAGCTGCTGTTCTGACGGCCCGCGCGTAAGGTGGCGGCGGC

       A  G  T  D  V  V  L  S  G  P  A  C  E  V  A  A

8201 CGGCGGCGGGCATCAGGACCGCAAGAAGCGCTGCGGCCTTACTCGGCCAC 8250
     ---------+---------+---------+---------+---------+
     GCCGCCGCCCGTAGTCCTGGCGTTCTTCGCGACGCCGGAATGAGCCGGTG

     A  A  A  P  M  L  V  A  L  L  A  A  A  K  S  P  W

8251 ATGGGCAAGATAGGACTGCTCGGCGCCGAGATCCTGCTGACCCTGCGCAT 8300
     ---------+---------+---------+---------+---------+
     TACCCGTTCTATCCTGACGAGCCGCGGCTCTAGGACGACTGGGACGCGTA

      M  P  L  I  P  S  S  P  A  S  I  R  S  V  R  R  M

8301 CCTCGTTCCGGTCATGCAGCGCCAGGTCCCATGCCGCGATCTGCGCGnnC 8350
     ---------+---------+---------+---------+---------+
     GGAGCAAGGCCAGTACGTCGCGGTCCAGGGTACGGCGCTAGACGCGCnnG

      R  T  G  T  M  <-- orf-16

8351 ATCAGCCCGCGCGGACCCTCGACGACGCGGAGGCAGATCGCCTCGCCGAT 8400
     ---------+---------+---------+---------+---------+
     TAGTCGGGCGCGCCTGGGAGCTGCTGCGCCTCCGTCTAGCGGAGCGGCTA
     F  L  M  G  T  G  A  P  L  E  K  R  F  M  L  G  Q
```

Fig. 7/15

```
                .         .         .         .         .
       CACGAGGTCCGAGAAGCCGGAATGACGGAGCACCTCGATATGGATGAACA
8401   ---------+---------+---------+---------+---------+   8450
       GTGCTCCAGGCTCTTCGGCCTTACTGCCTCGTGGAGCTATACCTACTTGT

                .         .         .         .         .
       CGTCCTCGGGGTGGCCGAAGATGTTGGCGAACCGGGAAAAGGCCCTTGGC
8451   ---------+---------+---------+---------+---------+   8500
       GCAGGAGCCCCACCGGCTTCTACAACCGCTTGGCCCTTTTCCGGGAACCG

                .         .         .         .         .
       CTTGTCGAACCACTTGACGCGGGCCGGACGCAGCGGCAnnCGTCCAGATG
8501   ---------+---------+---------+---------+---------+   8550
       GAACAGCTTGGTGAACTGCGCCCGGCCTGCGTCGCCGTnnGCAGGTCTAC

                .         .         .         .         .
       CTCGATCACCTCGGCATCCAGATCGGCGATnGGGGGGTGnCnGTCGCTTT
8551   ---------+---------+---------+---------+---------+   8600
       GAGCTAGTGGAGCCGTAGGTCTAGCCGCTAnCCCCCCACnGnCAGCGAAA

                .         .
       CnnnCGGTTCGATCGACAGGACCTC
8601   ---------+---------+-----   8625
       GnnnGCCAAGCTAGCTGTCCTGGAG
```

Fig. 8

```
  1  MTPKQQFPLR DLVEIRLAQI SGQFGVVSAP LGAAMSDAAL SPGKRFRAVL
 51  MLMVAESSGG VCDAMVDAAC AVEMVHAASL IFDDMPCMDD ARTRRGQPAT
101  HVAHGEGRAV LAGIALITEA MRILGEARGA TPDQRARLVA SMSRAMGPVG
151  LCAGQDLDLH APKDAAGIER EQDLKTGVLF VAGLEMLSII KGLDKAETEQ
201  LMAFGRQLGR VFQSYDDLLD VIGDKASTGK DTARDTAAPG PKGGLMAVGQ
251  MGDVAQHYRA SRAQLDELMR TRLFRGGQIA DLLARVLPHD IRRSA
```

Fig. 9

```
  1 MTDLTATSEA AIAQGSQSFA QAAKLMPPGI REDTVMLYAW CRHADDVIDG
 51 QVMGSAPEAG GDPQARLGAL RADTLAALHE DGPMSPPFAA LRQVARRHDF
101 PDLWPMDLIE GFAMDVADRE YRSLDDVLEY SYHVAGVVGV MMARVMGVQD
151 DAVLDRACDL GLAFQLTNIA RDVIDDAAIG RCYLPADWLA EAGATVEGPV
201 PSDALYSVII RLLDAAEPYY ASARQGLPHL PPRCAWSIAA ALRIYRAIGT
251 RIRQGGPEAY RQRISTSKAA KIGLLARGGL DAAASRLRGG EISRDGLWTR
301 PRA
```

Fig. 10

```
  1  MSSAIVIGAG FGGLALAIRL QSAGIATTIV EARDKPGGRA YVWNDQGHVF
 51  DAGPTVVTDP DSLRELWALS GQPMERDVTL LPVSPFYRLT WADGRSFEYV
101  NDDDELIRQV ASFNPADVDG YRRFHDYAEE VYREGYLKLG TTPFLKLGQM
151  LNAAPALMRL QAYRSVHSMV ARFIQDPHLR QAFSFHTLLV GGNPFSTSSI
201  YALIHALERR GGVWFAKGGT NQLVAGMVAL FERLGGTLLL NARVTRIDTE
251  GDRATGVTLL DGRQLRADTV ASNGDVMHSY RDLLGHTRRG RTKAAILNRQ
301  RWSMSLFVLH FGLSKRPENL AHHSVIFGPR YKGLVNEIFN GPRLPDDFSM
351  YLHSPCVTDP SLAPEGMSTH YVLAPVPHLG RADVDWEAEA PGYAERIFEE
401  LERRAIPDLR KHLTVSRIFS PADFSTELSA HHGSAFSVEP ILTQSAWFRP
451  HNRDRAIPNF YIVGAGTHPG AGIPGVVGSA KATAQVMLSD LAVA
```

Fig. 11

1 MSHDLLIAGA GLSGALIALA VRDRRPDARI VMLDARSGPS DQHTWSCHDT

51 DLSPEWLARL SPIRRGEWTD QEVAFPDHSR RLTTGYGSIE AGALIGLLQG

101 VDLRWNTHVA TLDDTGATLT DGSRIEAACV IDARGAVETP HLTVGFQKFV

151 GVEIETDAPH GVERPMIMDA TVPQMDGYRF IYLLPFSPTR ILIEDTRYSD

201 GGDLDDGALA QASLDYAARR GWTGQEMRRE RGILPIALAH DAIGFWRDHA

251 QGAVPVGLGA GLFHPVTGYS LPYAAQVADA IAARDLTTAS ARRAVRGWAI

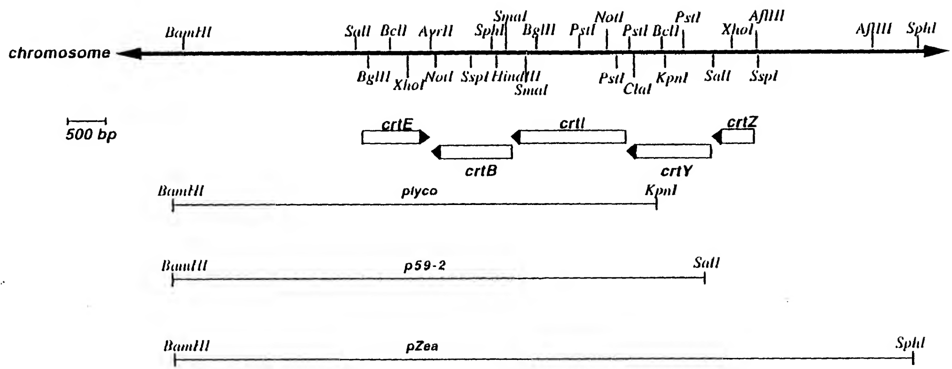
301 DRADRDRFLR LLNRMLFRGC PPDRRYRLLQ RFYRLPQPLI ERFYAGRLTL

351 ADRLRIVTGR PPIPLSQAVR CLPERPLLQE RA

Fig. 12

```
  1  MSTWAAILTV ILTVAAMELT AYSVHRWIMH GPLGWGWHKS HHDEDHDHAL
 51  EKNDLYGVIF AVISIVLFAI GAMGSDLAWW LAVGVTCYGL IYYFLHDGLV
101  HGRWPFRYVP KRGYLRRVYQ AHRMHHAVHG RENCVSFGFI WAPSVDSLKA
151  ELKRSGALLK DREGADRNT
```

Fig. 13



construct	crtE	crtB	crtI	crtY	crtZ	carotenoid
pLyco	+	+	+	-	-	LYCOPENE
p59-2	+	+	+	+	-	β-CAROTENE
pZea 4	+	+	+	+	+	ZEAXANTHIN

Fig. 14

crtE

#100: 5' tatatactagtaacaggagaaattacatATGACGCCCAAGCAGCAGCAATTC 3'

SpeI RBS NdeI

#101: 5' TATATACCCGGGTCAGCCGCGACGGCCTGTGG 3'

SmaI

crtZ

#104: 5' tatatgaattcaacaggagaaattacatATGAGCACTTGGGCCGCAATCC 3'

EcoRI RBS NdeI

#105: 5' GTTTCAGCTCTGCCTTGAGGC 3'

crtZ crtY

MUT1: 5' GCGAAGGGGCGGATCGCAATACgTGaaaggaggacacgtgATGAGCCATGATCTGCTGATCG 3'

PmlI

crtY crtI

MUT2: 5' GCCCCCTGCTGCAGGAGAGAGCtTGaaaggaggcaattgagATGAGTTCCGCCATCGTCATCG 3'

MunI

crtI crtB

MUT3: 5' GGTCATGCTGTCGGACCTGGCCGTCGCtTGaaaggaggatccaatcATGACCGATCTGACGGCGACTTCC 3'

BamHI

MUT5: 5' ATATATctcaattgcctcctttcaaGCTCTCTCCTGCAGCAGGG 3'

MunI crtY

MUT6: 5' atgattggatcctcctttcaaGCGACGGCCAGGTCCGACAGC 3'

BamHI crtI

CAR17 5' CAGAACCCATCACCTGCCCGTC 3'

cat3: 5' CGCGAATTCTCGCCGGCAATAGTTACC 3'

EcoRI

cat4: 5' GTCACATGCATGCATGTTACGAGCTCATAAGCATGTGACGTCTTCAACTAACGGGGCAGG 3'

SphI SacI AatII

Fig. 15

```
           HindIII         AflII      XbaI          EcoRI
CS1:  5'  AGCTTGGATCCTTAAGTACTCTAGAGTTTAAACG           3'
CS2:  3'         ACCTAGGAATTCATGAGATCTCAAATTTGCTTAA    5'
                         ScaI           PmeI

           SalI   AvrII          MluI          BamHI       HindIII
MUT7: 5'  TCGACCCTAGGCACGTGACGCGTCAATTGGATCCGCATGCAAGCTT           3'
MUT8: 3'         GGGATCCGTGCACTGCGCAGTTAACCTAGGCGTACGTTCGAACTAG    5'
                      PmlI             MunI         SphI

          1/2 PmlI                                                                                          SalI
MUT9:  5'  gtgtcctcctttcacGTATTGCGATCCGCCCCTTCGCGGTCCTTCAGCAGCGCGCCCGAGCGTTTCAGCTCTGCCTTGAGGCTG        3'
MUT10: 3'  cacaggaggaaagtgCATAACGCTAGGCGGGGAAGCGCCAGGAAGTCGTCGCGCGGGCTCGCAAAGTCGAGACGGAACTCCGACAGCT    5'
              RBS        crtZ

                                 SpeI
MUT11: 5'   TAAGAAACcctcctttA      3'
MUT12: 3'      TCTTTGggaggaaaTGATC  5'
                      RBS
```

Fig. 16

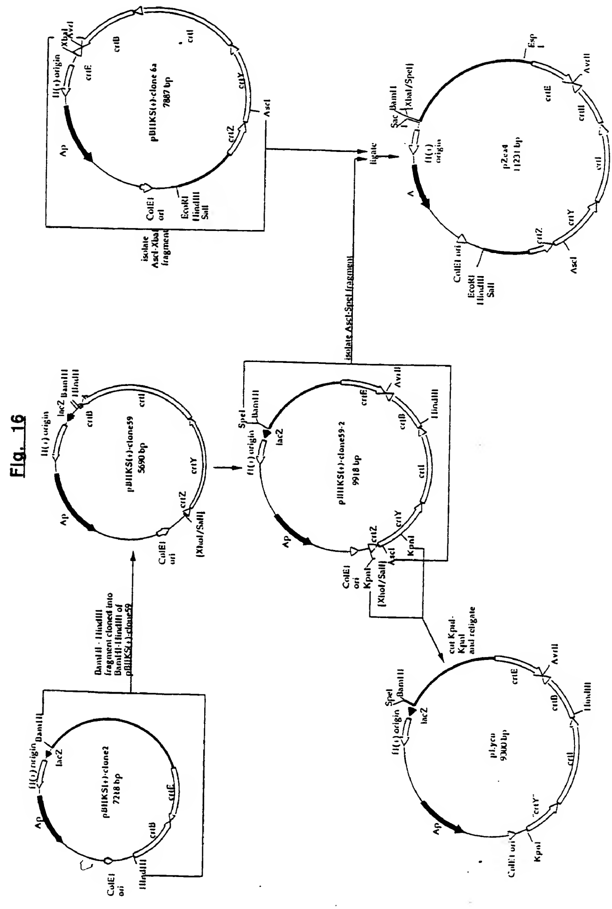

pBIIKS(+)-clone2
7218 bp
BamHI - HindIII
fragment cloned into
BamHI-HindIII of
pBIIKS(+)-clone59
pBIIKS(+)-clone59
5690 bp
pBIIKS(+)-clone59-2
9918 bp
cut KpnI-
KpnI
and religate
pLycn
9300 bp
pBIIKS(+)-clone 6a
7887 bp
isolate
AscI-XbaI
fragment
isolate AscI-SpeI fragment
ligate
pZea4
11231 bp
f1(+) origin
lacZ
Ap
ColEI ori
crtE
crtB
crtI
crtY
crtZ
BamHI
HindIII
SpeI
KpnI
AscI
AvrII
XbaI/SpeI
XhoI/SalI
EcoRI
SalI

Fig. 17

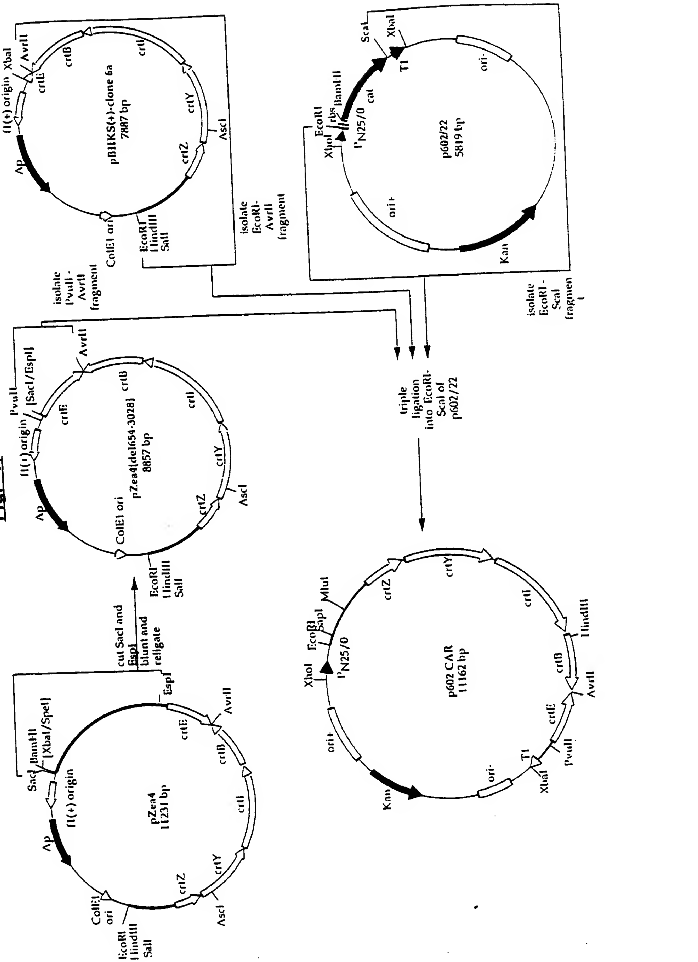
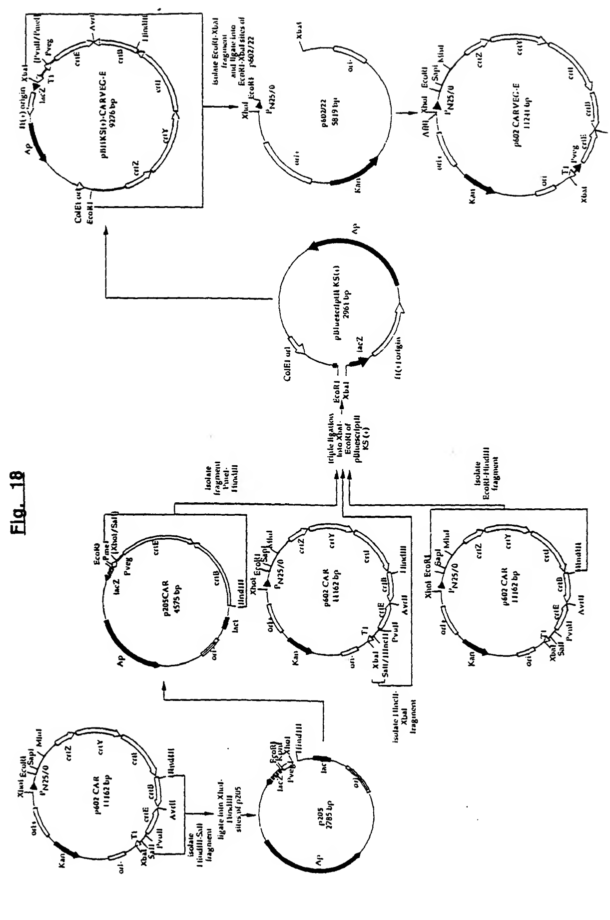

pBIIKS(+)-clone 6a
7887 bp
isolate
EcoRI-
AvrII
fragment
p602/22
5819 bp
isolate
EcoRI -
ScaI
fragment
isolate
PvuII -
AvrII
fragment
pZea4(del654-3028)
8857 bp
triple
ligation
into EcoRI-
ScaI of
p602/22
cut SacI and
EspI
blunt and
religate
pZea4
11231 bp
p602 CAR
11162 bp

Fig. 18


p602 CAR
11162 bp
p205
2785 bp
p205CAR
4575 bp
pBluescriptII KS(+)
2961 bp
pBIIKS(+)-CARVEC-E
9276 bp
p602/22
5819 bp
p602 CARVEC-E
11241 bp

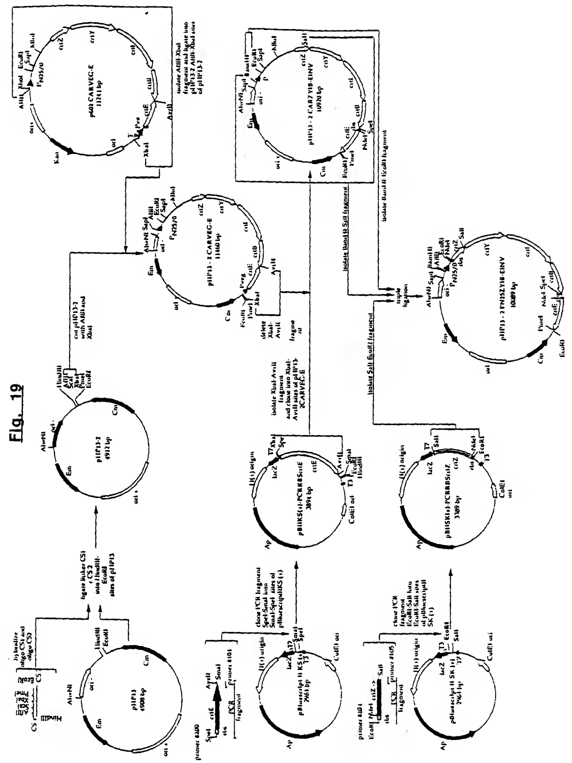


Fig. 19

Fig. 20/1

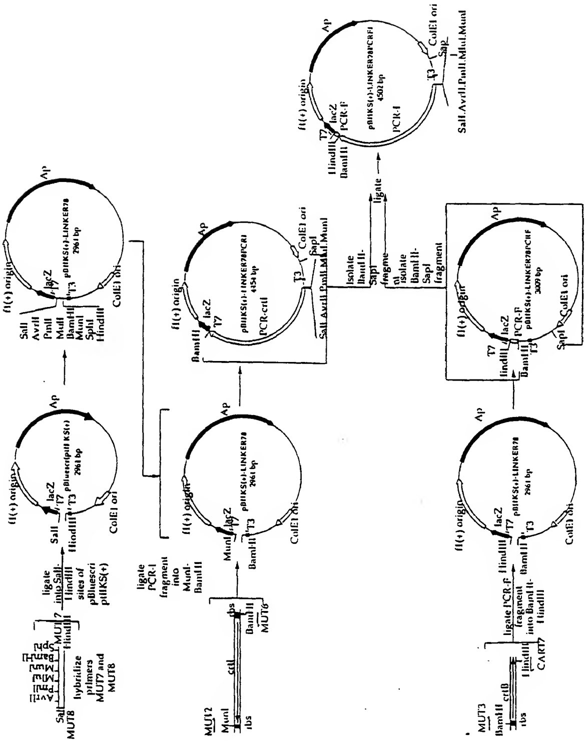

hybridize primers MUT7 and MUT8
ligate into SalI-HindIII sites of pBluescri pIIKS(+)
pBluescr(pIIKS(+)) 2961 bp
pBIIKS(+)-LINKER78 2961 bp
ligate PCR-I fragment into MunI-BamHI
pBIIKS(+)-LINKER78PCRI 4354 bp
ligate PCR-F fragment into BamHI-HindIII
pBIIKS(+)-LINKER78PCRF 3009 bp
isolate BamHI-SapI fragment
ligate
pBIIKS(+)-LINKER78PCRFI 4502 bp
f1(+) origin
lacZ
T7
T3
ColE1 ori
Ap
PCR-I
PCR-F
rbs
MUT2
MUT3
MUT6
CART7

Fig. 20/2

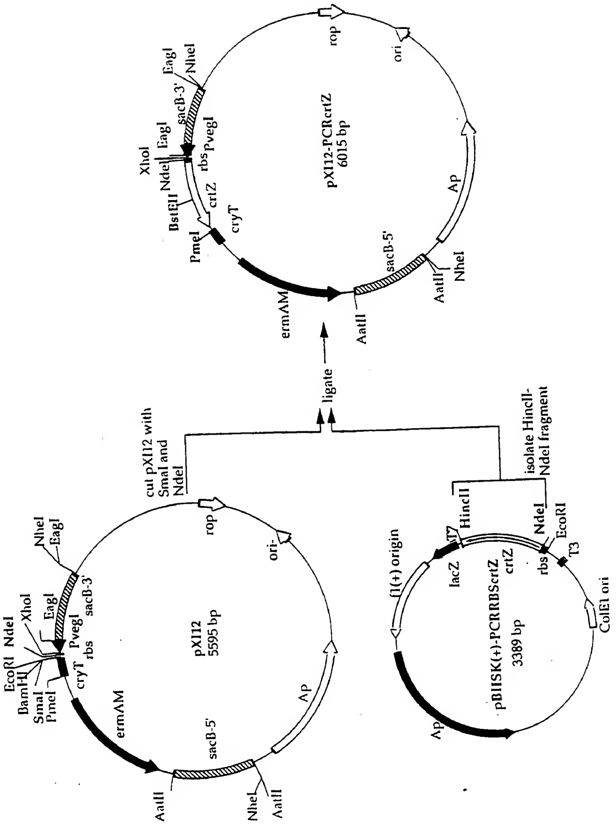

pXI12-PCRcrtZ
6015 bp
rop
ori
Ap
sacB-5'
NheI
AatII
AatII
ermAM
PmeI
cryT
BstEII
crtZ
NdeI
rbs
XhoI
EagI
PvegI
sacB-3'
EagI
NheI
ligate
cut pXI12 with
SmaI and
NdeI
isolate HincII-
NdeI fragment
pXI12
5595 bp
EcoRI
NdeI
XhoI
BamHI
SmaI
PmeI
EagI
PvegI
NheI
EagI
sacB-3'
cryT
rbs
rop
ori
ermAM
sacB-5'
Ap
AatII
NheI
AatII
pBIISK(+)-PCRRBScrtZ
3389 bp
f1(+) origin
lacZ
crtZ
rbs
HincII
NdeI
EcoRI
T3
ColE1 ori
Ap

Fig. 20/3

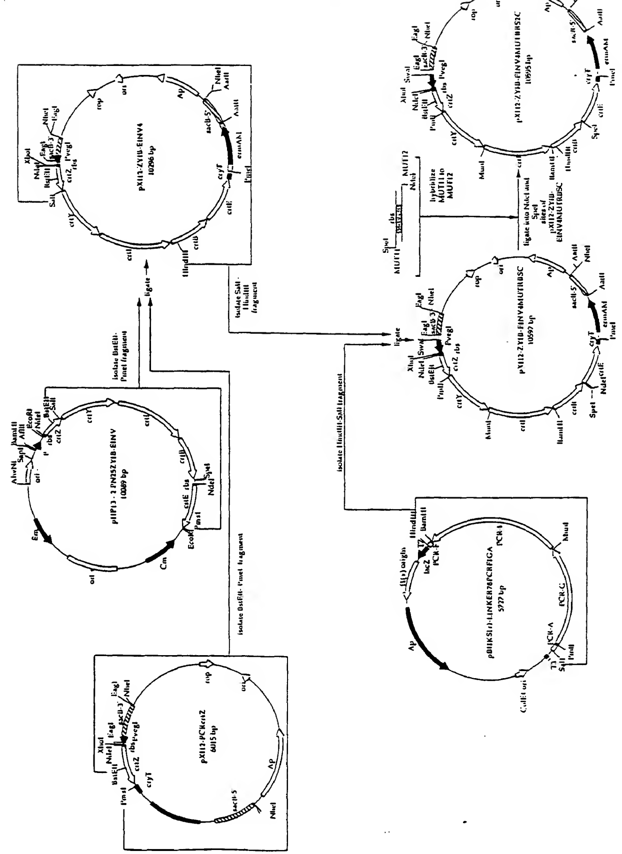


Fig. 20/4

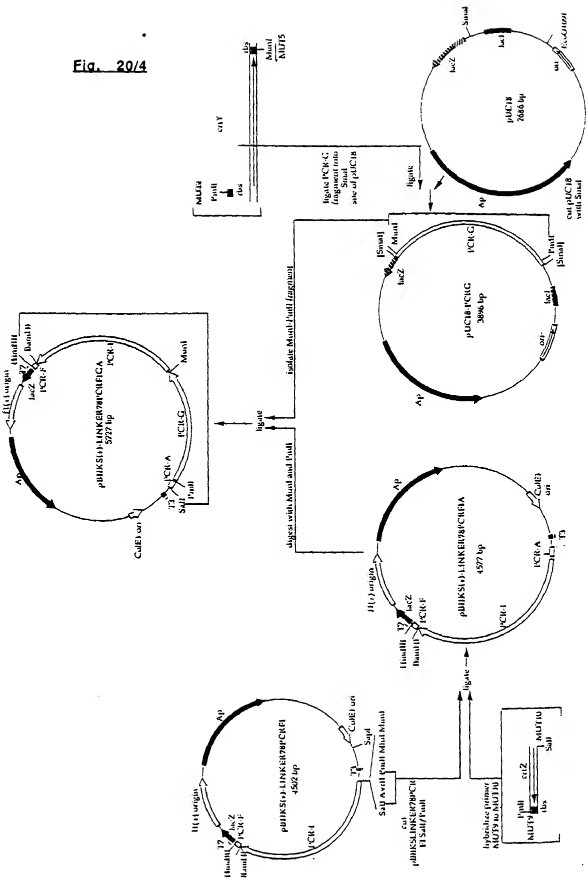

pUC18
2686 bp
pUC18-PCRG
3896 bp
pBIIKS(+)-LINKER78PCRFIGA
5727 bp
pBIIKS(+)-LINKER78PCRFIA
4577 bp
pBIIKS(+)-LINKER78PCRFI
4502 bp
isolate MluI-PstI fragment
digest with MluI and PstI
ligate

Fig. 21/1

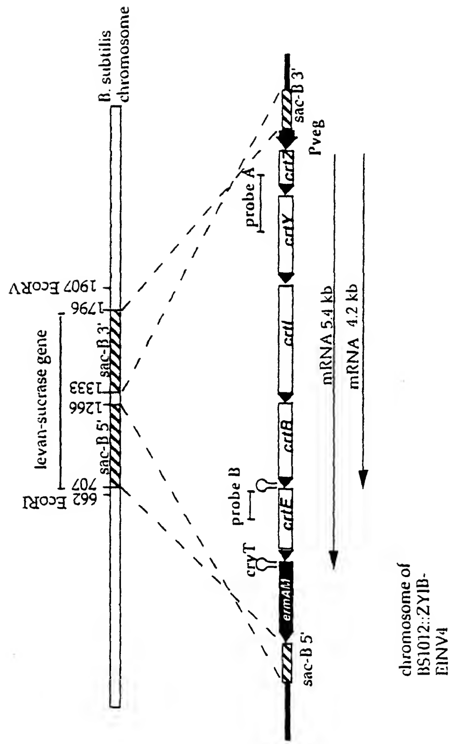

B. subtilis chromosome
1907 EcoRV
1796
sac-B 3'
1333
1266
sac-B 5'
707
662 EcoRI
levan-sucrase gene
sac-B 3'
Pveg
crtZ
probe A
crtY
crtI
crtB
probe B
crtE
cryT
ermAM
sac-B 5'
mRNA 5.4 kb
mRNA 4.2 kb
chromosome of BS1012::ZYIB-EINV4

Fig. 21/2

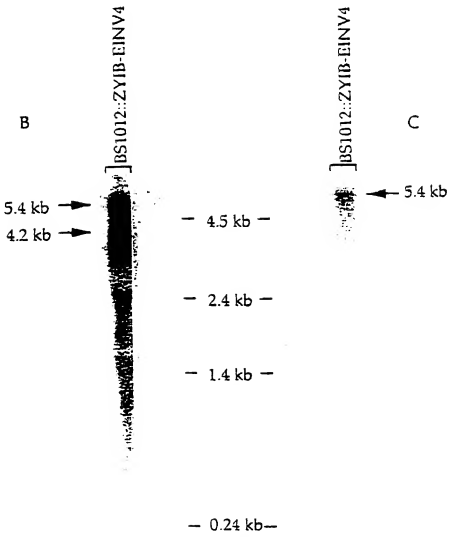

B
BS1012::ZYIB-EINV4
5.4 kb
4.2 kb
4.5 kb
2.4 kb
1.4 kb
0.24 kb
BS1012::ZYIB-EINV4
C
5.4 kb

Fig. 22

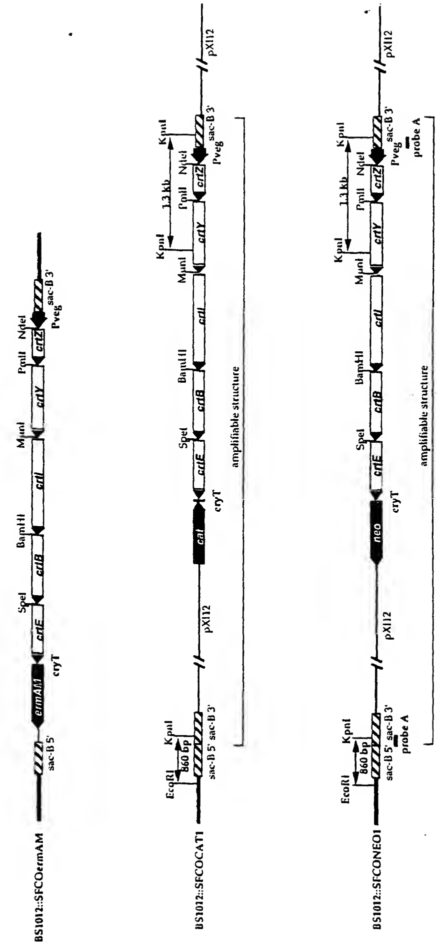

BS1012::SFCOermAM
sac-B 5'
ermAM
cryT
crtE
SpeI
crtB
BamHI
crtI
MunI
crtY
PmlI
crtZ
NdeI
Pveg
sac-B 3'
BS1012::SFCOCAT1
EcoRI
860 bp
KpnI
sac-B 5' sac-B 3'
pX112
cat
cryT
crtE
SpeI
crtB
BamHI
crtI
MunI
KpnI
crtY
PmlI
1.3 kb
crtZ
NdeI
KpnI
Pveg
sac-B 3'
pX112
amplifiable structure
BS1012::SFCONEO1
EcoRI
860 bp
KpnI
sac-B 5' sac-B 3'
probe A
pX112
neo
cryT
crtE
SpeI
crtB
BamHI
crtI
MunI
KpnI
crtY
PmlI
1.3 kb
crtZ
NdeI
KpnI
Pveg
sac-B 3'
probe A
pX112
amplifiable structure

Fig. 23

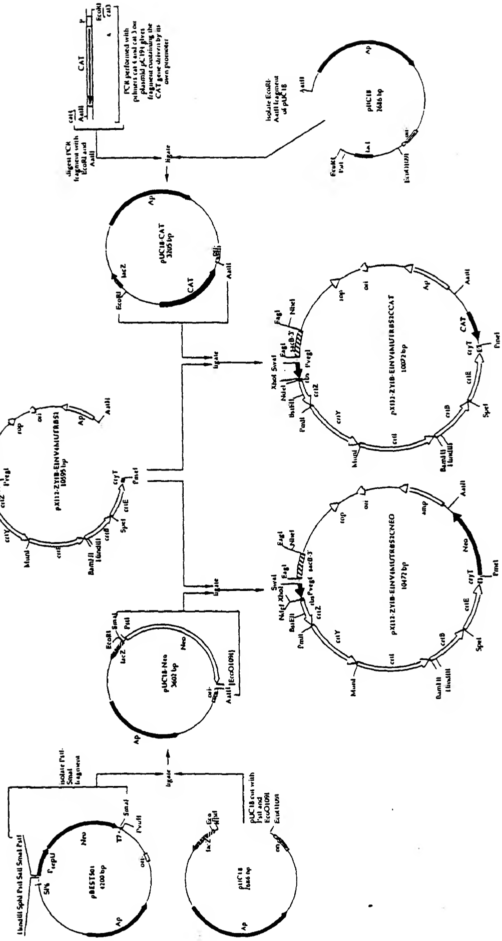


Fig. 24/1

```
     CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC
  1 ---------+---------+---------+---------+---------+---------+ 60
     GATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAG

     ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGA
 61 ---------+---------+---------+---------+---------+---------+ 120
     TAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCT

     GATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTC
121 ---------+---------+---------+---------+---------+---------+ 180
     CTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAG

     CAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACC
181 ---------+---------+---------+---------+---------+---------+ 240
     GTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGG

     CTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG
241 ---------+---------+---------+---------+---------+---------+ 300
     GATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGATTTCCCTC

     CCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAA
301 ---------+---------+---------+---------+---------+---------+ 360
     GGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTT

     AGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCAC
361 ---------+---------+---------+---------+---------+---------+ 420
     TCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTG

     CACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCG
421 ---------+---------+---------+---------+---------+---------+ 480
     GTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGGGTAAGCGGTAAGTCCGACGC

     CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
481 ---------+---------+---------+---------+---------+---------+ 540
     GTTGACAACCCTTCCCGCTAGCCACGCCCGGAGAAGCGATAATGCGGTCGACCGCTTTCC

     GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG
541 ---------+---------+---------+---------+---------+---------+ 600
     CCCTACACGACGTTCCGCTAATTCAACCCATTGCGGTCCCAAAAGGGTCAGTGCTGCAAC

     TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCA
601 ---------+---------+---------+---------+---------+---------+ 660
     ATTTTGCTGCCGGTCACTCGCGCGCATTATGCTGAGTGATATCCCGCTTAACCTCGAGGT

     CCGCGGTGGCGGCCGCTCTAGTGGATCCGCGCCTGGCCGTTCGCGATCAGCAGCCGCCCT
661 ---------+---------+---------+---------+---------+---------+ 720
     GGCGCCACCGCCGGCGAGATCACCTAGGCGCGGACCGGCAAGCGCTAGTCGTCGGCGGGA

     TGCGGATCGGTCAGCATCATCCCCATGAACCGCAGCGCACGACGCAGCGCGCGCCCCAGA
721 ---------+---------+---------+---------+---------+---------+ 780
     ACGCCTAGCCAGTCGTAGTAGGGGTACTTGGCGTCGCGTGCTGCGTCGCGCGCGGGGTCT

     TCGGGCGCGTCCAGCACGGCATGCGCCATCATCGCGAAGGCCCCCGGCGGCATGGGGCGC
781 ---------+---------+---------+---------+---------+---------+ 840
     AGCCCGCGCAGGTCGTGCCGTACGCGGTAGTAGCGCTTCCGGGGGCCGCCGTACCCCGCG

     GTGCCCATTCCGAAGAACTCGCAGCCTGTCCGCTGCGCAAGGTCGCGCCAGATCGCGCCG
841 ---------+---------+---------+---------+---------+---------+ 900
     CACGGGTAAGGCTTCTTGAGCGTCGGACAGGCGACGCGTTCCAGCGCGGTCTAGCGCGGC

     TATTCCGATGCAGTGACGGGCCCGATGCGCGTGGGCCCGCCCTGCCCCGCCGCCACCAGC
901 ---------+---------+---------+---------+---------+---------+ 960
     ATAAGGCTACGTCACTGCCCGGGCTACGCGCACCCGGGCGGGACGGGGCGGCGGTGGTCG
```

Fig. 24/2

```
     GCATCGCGCACGAACCCTTCCGAGATGATGTGCTGATCCATGGCCCGTCATTGCAAAACC
961  ---------+---------+---------+---------+---------+---------+ 1020
     CGTAGCGCGTGCTTGGGAAGGCTCTACTACACGACTAGGTACCGGGCAGTAACGTTTTGG

     GATCACCGATCCTGTCGCGTGATGGCATTGTTTGCAATGCCCCGAGGGCTAGGATGGCGC
1021 ---------+---------+---------+---------+---------+---------+ 1080
     CTAGTGGCTAGGACAGCGCACTACCGTAACAAACGTTACGGGGCTCCCGATCCTACCGCG

     GAAGGATCAAGGGGGGGAGAGACATGGAAATCGAGGGACGGGTCTTTGTCGTCACGGGCG
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CTTCCTAGTTCCCCCCCTCTCTGTACCTTTAGCTCCCTGCCCAGAAACAGCAGTGCCCGC

     CCGCATCGGGTCTGGGGGCGGCCTCGGCGCGGATGCTGGCCCAAGGCGGCGCGAAGGTCG
1141 ---------+---------+---------+---------+---------+---------+ 1200
     GGCGTAGCCCAGACCCCCGCCGGAGCCGCGCCTACGACCGGGTTCCGCCGCGCTTCCAGC

     TGCTGGCCGATCTGGCGGAACCGAAGGACGCGCCCGAAGGCGCGGTTCACGCGGCCTGCG
1201 ---------+---------+---------+---------+---------+---------+ 1260
     ACGACCGGCTAGACCGCCTTGGCTTCCTGCGCGGGCTTCCGCGCCAAGTGCGCCGGACGC

     ACGTGACCGACGCGACCGCTGCGCAGACGGCCATCGCGCTGGCGACCGACCGCTTCGGCA
1261 ---------+---------+---------+---------+---------+---------+ 1320
     TGCACTGGCTGCGCTGGCGACGCGTCTGCCGGTAGCGCGACCGCTGGCTGGCGAAGCCGT

     GGCTGGACGGCCTTGTGAACTGCGCGGGCATCGCGCCGGCCGAACGGATGCTGGGCCGCG
1321 ---------+---------+---------+---------+---------+---------+ 1380
     CCGACCTGCCGGAACACTTGACGCGCCCGTAGCGCGGCCGGCTTGCCTACGACCCGGCGC

     ACGGGCCGCATGGACTGGACAGCTTTGCCCGTGCGGTCACGATCAACCTGATCGGCAGCT
1381 ---------+---------+---------+---------+---------+---------+ 1440
     TGCCCGGCGTACCTGACCTGTCGAAACGGGCACGCCAGTGCTAGTTGGACTAGCCGTCGA

     TCAACATGGCCCGCCTTGCAGCCGAGGCGATGGCCCGGAACGAGCCCGTCCGGGGCGAGC
1441 ---------+---------+---------+---------+---------+---------+ 1500
     AGTTGTACCGGGCGGAACGTCGGCTCCGCTACCGGGCCTTGCTCGGGCAGGCCCCGCTCG

     GTGGCGTGATCGTCAACACGGCCTCGATCGCGGCGCAGGACGGACAGATCGGACAGGTCG
1501 ---------+---------+---------+---------+---------+---------+ 1560
     CACCGCACTAGCAGTTGTGCCGGAGCTAGCGCCGCGTCCTGCCTGTCTAGCCTGTCCAGC

     CCTATGCGGCCAGCAAGGCGGGCGTGGCGGGCATGACGCTGCCGATGGCCCGCGACCTTG
1561 ---------+---------+---------+---------+---------+---------+ 1620
     GGATACGCCGGTCGTTCCGCCCGCACCGCCCGTACTGCGACGGCTACCGGGCGCTGGAAC

     CGCGGCACGGCATCCGCGTCATGACCATCGCGCCCGGCATCTTCCGCACCCCGATGCTGG
1621 ---------+---------+---------+---------+---------+---------+ 1680
     GCGCCGTGCCGTAGGCGCAGTACTGGTAGCGCGGGCCGTAGAAGGCGTGGGGCTACGACC

     AGGGGCTGCCGCAGGACGTTCAGGACAGCCTGGGCGCGGCGGTGCCCTTCCCCTCGCGGC
1681 ---------+---------+---------+---------+---------+---------+ 1740
     TCCCCGACGGCGTCCTGCAAGTCCTGTCGGACCCGCGCCGCCACGGGAAGGGGAGCGCCG

     TGGGAGAGCCGTCGGAATACGCGGCGCTGTTGCACCACATCATCGCGAACCCCATGCTGA
1741 ---------+---------+---------+---------+---------+---------+ 1800
     ACCCTCTCGGCAGCCTTATGCGCCGCGACAACGTGGTGTAGTAGCGCTTGGGGTACGACT

     ACGGAGAGGTCATCCGCCTCGACGGCGCATTGCGCATGGCCCCCAAGTGAAGGAGCGTTT
1801 ---------+---------+---------+---------+---------+---------+ 1860
     TGCCTCTCCAGTAGGCGGAGCTGCCGCGTAACGCGTACCGGGGGTTCACTTCCTCGCAAA

     CATGGACCCCATCGTCATCACCGGCGCGATGCGCACCCCGATGGGGGCATTCCAGGGCGA
1861 ---------+---------+---------+---------+---------+---------+ 1920
     GTACCTGGGGTAGCAGTAGTGGCCGCGCTACGCGTGGGGCTACCCCCGTAAGGTCCCGCT

     TCTTGCCGCGATGGATGCCCCGACCCTTGGCGCGGACGCGATCCGCGCCGCGCTGAACGG
1921 ---------+---------+---------+---------+---------+---------+ 1980
     AGAACGGCGCTACCTACGGGGCTGGGAACCGCGCCTGCGCTAGGCGCGGCGCGACTTGCC
```

Fig. 24/3

```
     CCTGTCGCCCGACATGGTGGACGAGGTGCTGATGGGCTGCGTCCTCGCCGCGGGCCAGGG
1981 ---------+---------+---------+---------+---------+---------+ 2040
     GGACAGCGGGCTGTACCACCTGCTCCACGACTACCCGACGCAGGAGCGGCGCCCGGTCCC

     TCAGGCACCGGCACGTCAGGCGGCGCTTGGCGCCGGACTGCCGCTGTCGACGGGCACGAC
2041 ---------+---------+---------+---------+---------+---------+ 2100
     AGTCCGTGGCCGTGCAGTCCGCCGCGAACCGCGGCCTGACGGCGACAGCTGCCCGTGCTG

     CACCATCAACGAGATGTGCGGATCGGGCATGAAGGCCGCGATGCTGGGCCATGACCTGAT
2101 ---------+---------+---------+---------+---------+---------+ 2160
     GTGGTAGTTGCTCTACACGCCTAGCCCGTACTTCCGGCGCTACGACCCGGTACTGGACTA

     CGCCGCGGGATCGGCGGGCATCGTCGTCGCCGGCGGGATGGAGAGCATGTCGAACGCCCC
2161 ---------+---------+---------+---------+---------+---------+ 2220
     GCGGCGCCCTAGCCGCCCGTAGCAGCAGCGGCCGCCCTACCTCTCGTACAGCTTGCGGGG

     CTACCTGCTGCCCAAGGCGCGGTCGGGGATGCGCATGGGCCATGACCGTGTGCTGGATCA
2221 ---------+---------+---------+---------+---------+---------+ 2280
     GATGGACGACGGGTTCCGCGCCAGCCCCTACGCGTACCCGGTACTGGCACACGACCTAGT

     CATGTTCCTCGACGGGTTGGAGGACGCCTATGACAAGGGCCGCCTGATGGGCACCTTCGC
2281 ---------+---------+---------+---------+---------+---------+ 2340
     GTACAAGGAGCTGCCCAACCTCCTGCGGATACTGTTCCCGGCGGACTACCCGTGGAAGCG

     CGAGGATTGCGCCGGCGATCACGGTTTCACCCGCGAGGCGCAGGACGACTATGCGCTGAC
2341 ---------+---------+---------+---------+---------+---------+ 2400
     GCTCCTAACGCGGCCGCTAGTGCCAAAGTGGGCGCTCCGCGTCCTGCTGATACGCGACTG

     CAGCCTGGCCCGCGCGCAGGACGCCATCGCCAGCGGTGCCTTCGCCGCCGAGATCGCGCC
2401 ---------+---------+---------+---------+---------+---------+ 2460
     GTCGGACCGGGCGCGCGTCCTGCGGTAGCGGTCGCCACGGAAGCGGCGGCTCTAGCGCGG

     CGTGACCGTCACGGCACGCAAGGTGCAGACCACCGTCGATACCGACGAGATGCCCGGCAA
2461 ---------+---------+---------+---------+---------+---------+ 2520
     GCACTGGCAGTGCCGTGCGTTCCACGTCTGGTGGCAGCTATGGCTGCTCTACGGGCCGTT

     GGCCCGCCCCGAGAAGATCCCCCATCTGAAGCCCGCCTTCCGTGACGGTGGCACGGTCAC
2521 ---------+---------+---------+---------+---------+---------+ 2580
     CCGGGCGGGGCTCTTCTAGGGGGTAGACTTCGGGCGGAAGGCACTGCCACCGTGCCAGTG

     GGCGGCGAACAGCTCGTCGATCTCGGACGGGGCGGCGGCGCTGGTGATGATGCGCCAGTC
2581 ---------+---------+---------+---------+---------+---------+ 2640
     CCGCCGCTTGTCGAGCAGCTAGAGCCTGCCCCGCCGCCGCGACCACTACTACGCGGTCAG

     GCAGGCCGAGAAGCTGGGCCTGACGCCGATCGCGCGGATCATCGGTCATGCGACCCATGC
2641 ---------+---------+---------+---------+---------+---------+ 2700
     CGTCCGGCTCTTCGACCCGGACTGCGGCTAGCGCGCCTAGTAGCCAGTACGCTGGGTACG

     CGACCGTCCCGGCCTGTTCCCGACGGCCCCCATCGGCGCGATGCGCAAGCTGCTGGACCG
2701 ---------+---------+---------+---------+---------+---------+ 2760
     GCTGGCAGGGCCGGACAAGGGCTGCCGGGGGTAGCCGCGCTACGCGTTCGACGACCTGGC

     CACGGACACCCGCCTTGGCGATTACGACCTGTTCGAGGTGAACGAGGCATTCGCCGTCGT
2761 ---------+---------+---------+---------+---------+---------+ 2820
     GTGCCTGTGGGCGGAACCGCTAATGCTGGACAAGCTCCACTTGCTCCGTAAGCGGCAGCA

     CGCCATGATCGCGATGAAGGAGCTTGGCCTGCCACACGATGCCACGAACATCAACGGCGG
2821 ---------+---------+---------+---------+---------+---------+ 2880
     GCGGTACTAGCGCTACTTCCTCGAACCGGACGGTGTGCTACGGTGCTTGTAGTTGCCGCC

     GGCCTGCGCGCTTGGGCATCCCATCGGCGCGTCGGGGGCGCGGATCATGGTCACGCTGCT
2881 ---------+---------+---------+---------+---------+---------+ 2940
     CCGGACGCGCGAACCCGTAGGGTAGCCGCGCAGCCCCCGCGCCTAGTACCAGTGCGACGA

     GAACGCGATGGCGGCGCGGGGCGCGACGCGCGGGGCCGCATCCGTCTGCATCGGCGGGGG
2941 ---------+---------+---------+---------+---------+---------+ 3000
     CTTGCGCTACCGCCGCGCCCCGCGCTGCGCGCCCCGGCGTAGGCAGACGTAGCCGCCCCC
```

Fig. 24/4

```
     CGAGGCGACGGCCATCGCGCTGGAACGGCTGAGCTAATTCATTTGCGCGAATCCGCGTTT
3001 ---------+---------+---------+---------+---------+---------+ 3060
     GCTCCGCTGCCGGTAGCGCGACCTTGCCGACTCGATTAAGTAAACGCGCTTAGGCGCAAA

     TTCGTGCACGATGGGGGAACCGGAAACGGCCACGCCTGTTGTGGTTGCGTCGACCTGTCT
3061 ---------+---------+---------+---------+---------+---------+ 3120
     AAGCACGTGCTACCCCCTTGGCCTTTGCCGGTGCGGACAACACCAACGCAGCTGGACAGA

     TCGGGCCATGCCCGTGACGCGATGTGGCAGGCGCATGGGGCGTTGCCGATCCGGTCGCAT
3121 ---------+---------+---------+---------+---------+---------+ 3180
     AGCCCGGTACGGGCACTGCGCTACACCGTCCGCGTACCCCGCAACGGCTAGGCCAGCGTA

     GACTGACGCAACGAAGGCACCGATGACGCCCAAGCAGCAATTCCCCCTACGCGATCTGGT
3181 ---------+---------+---------+---------+---------+---------+ 3240
     CTGACTGCGTTGCTTCCGTGGCTACTGCGGGTTCGTCGTTAAGGGGGATGCGCTAGACCA

     CGAGATCAGGCTGGCGCAGATCTCGGGCCAGTTCGGCGTGGTCTCGGCCCCGCTCGGCGC
3241 ---------+---------+---------+---------+---------+---------+ 3300
     GCTCTAGTCCGACCGCGTCTAGAGCCCGGTCAAGCCGCACCAGAGCCGGGGCGAGCCGCG

     GGCCATGAGCGATGCCGCCCTGTCCCCCGGCAAACGCTTTCGCGCCGTGCTGATGCTGAT
3301 ---------+---------+---------+---------+---------+---------+ 3360
     CCGGTACTCGCTACGGCGGGACAGGGGGCCGTTTGCGAAAGCGCGGCACGACTACGACTA

     GGTCGCCGAAAGCTCGGGCGGGTCTGCGATGCGATGGTCGATGCCGCCTGCGCGGTCGA
3361 ---------+---------+---------+---------+---------+---------+ 3420
     CCAGCGGCTTTCGAGCCCGCCCAGACGCTACGCTACCAGCTACGGCGGACGCGCCAGCT

     GATGGTCCATGCCGCATCGCTGATCTTCGACGACATGCCCTGCATGGACGATGCCAGGAC
3421 ---------+---------+---------+---------+---------+---------+ 3480
     CTACCAGGTACGGCGTAGCGACTAGAAGCTGCTGTACGGGACGTACCTGCTACGGTCCTG

     CCGTCGCGGTCAGCCCGCCACCCATGTCGCCCATGGCGAGGGGCGCGCGGTGCTTGCGGG
3481 ---------+---------+---------+---------+---------+---------+ 3540
     GGCAGCGCCAGTCGGGCGGTGGGTACAGCGGGTACCGCTCCCCGCGCGCCACGAACGCCC

     CATCGCCCTGATCACCGAGGCCATGCGGATTTTGGGCGAGGCGCGCGGCGCGACGCCGGA
3541 ---------+---------+---------+---------+---------+---------+ 3600
     GTAGCGGGACTAGTGGCTCCGGTACGCCTAAAACCCGCTCCGCGCGCCGCGCTGCGGCCT

     TCAGCGCGCAAGGCTGGTCGCATCCATGTCGCGCGCGATGGGACCGGTGGGGCTGTGCGC
3601 ---------+---------+---------+---------+---------+---------+ 3660
     AGTCGCGCGTTCCGACCAGCGTAGGTACAGCGCGCGCTACCCTGGCCACCCCGACACGCG

     AGGGCAGGATCTGGACCTGCACGCCCCCAAGGACGCCGCCGGGATCGAACGTGAACAGGA
3661 ---------+---------+---------+---------+---------+---------+ 3720
     TCCCGTCCTAGACCTGGACGTGCGGGGGTTCCTGCGGCGGCCCTAGCTTGCACTTGTCCT

     CCTCAAGACCGGCGTGCTGTTCGTCGCGGGCCTCGAGATGCTGTCCATTATTAAGGGTCT
3721 ---------+---------+---------+---------+---------+---------+ 3780
     GGAGTTCTGGCCGCACGACAAGCAGCGCCCGGAGCTCTACGACAGGTAATAATTCCCAGA

     GGACAAGGCCGAGACCGAGCAGCTCATGGCCTTCGGGCGTCAGCTTGGTCGGGTCTTCCA
3781 ---------+---------+---------+---------+---------+---------+ 3840
     CCTGTTCCGGCTCTGGCTCGTCGAGTACCGGAAGCCCGCAGTCGAACCAGCCCAGAAGGT

     GTCCTATGACGACCTGCTGGACGTGATCGGCGACAAGGCCAGCACCGGCAAGGATACGGC
3841 ---------+---------+---------+---------+---------+---------+ 3900
     CAGGATACTGCTGGACGACCTGCACTAGCCGCTGTTCCGGTCGTGGCCGTTCCTATGCCG

     GCGCGACACCGCCGCCCCCGGCCCAAAGGCGGCCTGATGGCGGTCGGACAGATGGGCGA
3901 ---------+---------+---------+---------+---------+---------+ 3960
     CGCGCTGTGGCGGCGGGGGCCGGGTTTCCCGCCGGACTACCGCCAGCCTGTCTACCCGCT

     CGTGGCGCAGCATTACCGCGCCAGCCGCGCGCAACTGGACGAGCTGATGCGCACCCGGCT
3961 ---------+---------+---------+---------+---------+---------+ 4020
     GCACCGCGTCGTAATGGCGCGGTCGGCGCGCGTTGACCTGCTCGACTACGCGTGGGCCGA
```

Fig. 24/5

```
     GTTCCGCGGGGGGCAGATCGCGGACCTGCTGGCCCGCGTGCTGCCGCATGACATCCGCCG
4021 ---------+---------+---------+---------+---------+---------+ 4080
     CAAGGCGCCCCCCGTCTAGCGCCTGGACGACCGGGCGCACGACGGCGTACTGTAGGCGGC

     CAGCGCCTAGGCGCGCGGTCGGGTCCACAGGCCGTCGCGGCTGATTTCGCCGCCGCGCAG
4081 ---------+---------+---------+---------+---------+---------+ 4140
     GTCGCGGATCCGCGCGCCAGCCCAGGTGTCCGGCAGCGCCGACTAAAGCGGCGGCGCGTC

     GCGCGATGCGGCCGCGTCCAAGCCTCCGCGCGCCAGAAGCCCGATCTTGGCAGCCTTCGA
4141 ---------+---------+---------+---------+---------+---------+ 4200
     CGCGCTACGCCGGCGCAGGTTCGGAGGCGCGCGGTCTTCGGGCTAGAACCGTCGGAAGCT

     CGTGCTGATCCGCTGGCGATAGGCCTCGGGGCCACCCTGCCGGATGCGCGTCCCGATTGC
4201 ---------+---------+---------+---------+---------+---------+ 4260
     GCACGACTAGGCGACCGCTATCCGGAGCCCCGGTGGGACGGCCTACGCGCAGGGCTAACG

     GCGATAGATACGCAGCGCGGCGGCGATCGACCACGCGCAGCGCGGCGGCAGATGCGGAAG
4261 ---------+---------+---------+---------+---------+---------+ 4320
     CGCTATCTATGCGTCGCGCCGCCGCTAGCTGGTGCGCGTCGCGCCGCCGTCTACGCCTTC

     CCCCTGCCGCGCCGAGGCATAATAGGGCTCGGCCGCGTCAAGCAGGCGGATGATGACGGA
4321 ---------+---------+---------+---------+---------+---------+ 4380
     GGGGACGGCGCGGCTCCGTATTATCCCGAGCCGGCGCAGTTCGTCCGCCTACTACTGCCT

     ATAGAGCGCGTCCGAAGGCACCGGACCCTCAACCGTCGCCCCCGCCTCGGCCAGCCAGTC
4381 ---------+---------+---------+---------+---------+---------+ 4440
     TATCTCGCGCAGGCTTCCGTGGCCTGGGAGTTGGCAGCGGGGGCGGAGCCGGTCGGTCAG

     GGCAGGCAGATAGCAGCGCCCGATGGCGGCATCGTCGATCACGTCGCGAGCGATGTTCGT
4441 ---------+---------+---------+---------+---------+---------+ 4500
     CCGTCCGTCTATCGTCGCGGGCTACCGCCGTAGCAGCTAGTGCAGCGCTCGCTACAAGCA

     CAGCTGGAACGCAAGGCCCAGATCGCAGGCGCGATCCAGCACCGCATCGTCCTGCACGCC
4501 ---------+---------+---------+---------+---------+---------+ 4560
     GTCGACCTTGCGTTCCGGGTCTAGCGTCCGCGCTAGGTCGTGGCGTAGCAGGACGTGCGG

     CATCACCCGCGCCATCATCACGCCCACGACCCCCGCGACGTGGTAGGAATATTCCAGCAC
4561 ---------+---------+---------+---------+---------+---------+ 4620
     GTAGTGGGCGCGGTAGTAGTGCGGGTGCTGGGGGCGCTGCACCATCCTTATAAGGTCGTG

     GTCATCCAGGCTGCGGTATTCGCGATCCGCGACATCCATCGCGAAACCCTCGATCAGGTC
4621 ---------+---------+---------+---------+---------+---------+ 4680
     CAGTAGGTCCGACGCCATAAGCGCTAGGCGCTGTAGGTAGCGCTTTGGGAGCTAGTCCAG

     CATCGGCCAAAGGTCCGGGAAATCATGCCGCCGGGCGACCTGGCGCAGCGCCGCGAAGGG
4681 ---------+---------+---------+---------+---------+---------+ 4740
     GTAGCCGGTTTCCAGGCCCTTTAGTACGGCGGCCCGCTGGACCGCGTCGCGGCGCTTCCC

     CGGCGACATCGGGCCGTCCTCGTGCAGCGCGGCCAGCGTGTCGGCGCGCAGCGCCCCCAG
4741 ---------+---------+---------+---------+---------+---------+ 4800
     GCCGCTGTAGCCCGGCAGGAGCACGTCGCGCCGGTCGCACAGCCGCGCGTCGCGGGGGTC

     CCGCGCCTGTGGGTCGCCGCCCGCCTCGGGGGCAGAACCCATCACCTGCCCGTCGATCAC
4801 ---------+---------+---------+---------+---------+---------+ 4860
     GGCGCGGACACCCAGCGGCGGGCGGAGCCCCCGTCTTGGGTAGTGGACGGGCAGCTAGTG

     GTCATCCGCATGCCTGCACCAGGCATAGAGCATGACCGTATCCTCGCGGATGCCGGGCGG
4861 ---------+---------+---------+---------+---------+---------+ 4920
     CAGTAGGCGTACGGACGTGGTCCGTATCTCGTACTGGCATAGGAGCGCCTACGGCCCGCC

     CATCAGCTTGGCCGCCTGCGCGAAGCTTTGCGAACCCTGCGCGATGGCCGCTTCGGAAGT
4921 ---------+---------+---------+---------+---------+---------+ 4980
     GTAGTCGAACCGGCGGACGCGCTTCGAAACGCTTGGGACGCGCTACCGGCGAAGCCTTCA

     CGCCGTCAGATCGGTCATGCGACGGCCAGGTCCGACAGCATGACCTGCGCCGTGGCCTTG
4981 ---------+---------+---------+---------+---------+---------+ 5040
     GCGGCAGTCTAGCCAGTACGCTGCCGGTCCAGGCTGTCGTACTGGACGCGGCACCGGAAC
```

Fig. 24/6

```
     GCGCTGCCAACGACACCCGGGATGCCCGCACCCGGATGCGTGCCCGCCCCCACGATGTAG
5041 ---------+---------+---------+---------+---------+---------+ 5100
     CGCGACGGTTGCTGTGGGCCCTACGGGCGTGGGCCTACGCACGGGCGGGGGTGCTACATC

     AAGTTCGGGATCGCGCGGTCGCGGTTATGCGGGCGGAACCAGGCGGATTGCGTCAGGATC
5101 ---------+---------+---------+---------+---------+---------+ 5160
     TTCAAGCCCTAGCGCGCCAGCGCCAATACGCCCGCCTTGGTCCGCCTAACGCAGTCCTAG

     GGCTCGACCGAGAAGGCGCTGCCGTGATGGGCCGACAGTTCGGTGCTGAAATCGGCGGGG
5161 ---------+---------+---------+---------+---------+---------+ 5220
     CCGAGCTGGCTCTTCCGCGACGGCACTACCCGGCTGTCAAGCCACGACTTTAGCCGCCCC

     CTGAAGATGCGGCTGACGGTCAGGTGCTTGCGCAGGTCGGGGATGGCGCGGCGCTCCAGT
5221 ---------+---------+---------+---------+---------+---------+ 5280
     GACTTCTACGCCGACTGCCAGTCCACGAACGCGTCCAGCCCCTACCGCGCCGCGAGGTCA

     TCCTCGAAGATGCGCTCGGCATAGCCCGGGGCCTCGGCTTCCCAATCGACATCGGCGCGG
5281 ---------+---------+---------+---------+---------+---------+ 5340
     AGGAGCTTCTACGCGAGCCGTATCGGGCCCCGGAGCCGAAGGGTTAGCTGTAGCCGCGCC

     CCCAGATGCGGAACGGGCGCAAGGACGTAATGCGTGGACATCCCCTCGGGGGCCAGGCTG
5341 ---------+---------+---------+---------+---------+---------+ 5400
     GGGTCTACGCCTTGCCCGCGTTCCTGCATTACGCACCTGTAGGGGAGCCCCCGGTCCGAC

     GGATCGGTCACGCAGGGCGAATGCAGATACATCGAGAAATCGTCCGGCAGGCGTGGCCCG
5401 ---------+---------+---------+---------+---------+---------+ 5460
     CCTAGCCAGTGCGTCCCGCTTACGTCTATGTAGCTCTTTAGCAGGCCGTCCGCACCGGGC

     TTGAAGATCTCGTTCACCAGCCCCTTGTAGCGCGGGCCGAAGATGACGCTGTGGTGGGCC
5461 ---------+---------+---------+---------+---------+---------+ 5520
     AACTTCTAGAGCAAGTGGTCGGGGAACATCGCGCCCGGCTTCTACTGCGACACCACCCGG

     AGGTTCTCGGGGCGCTTGGACAGGCCGAAATGCAGCACGAACAGCGACATCGACCAGCGC
5521 ---------+---------+---------+---------+---------+---------+ 5580
     TCCAAGAGCCCCGCGAACCTGTCCGGCTTTACGTCGTGCTTGTCGCTGTAGCTGGTCGCG

     TGCCGGTTCAGGATCGCGGCCTTGGTGCGCCCGCGGCGGGTATGGCCCAGCAGGTCGCGA
5581 ---------+---------+---------+---------+---------+---------+ 5640
     ACGGCCAAGTCCTAGCGCCGGAACCACGCGGGCGCCGCCCATACCGGGTCGTCCAGCGCT

     TAGCTGTGCATCACGTCGCCGTTGCTGGCCACCGTATCCGCGCGCAACTGCCGCCCGTCC
5641 ---------+---------+---------+---------+---------+---------+ 5700
     ATCGACACGTAGTGCAGCGGCAACGACCGGTGGCATAGGCGCGCGTTGACGGCGGGCAGG

     AGCAGCGTGACGCCCGTGGCGCGATCGCCCTCGGTGTCGATCCGCGTGACGCGGGCATTC
5701 ---------+---------+---------+---------+---------+---------+ 5760
     TCGTCGCACTGCGGGCACCGCGCTAGCGGGAGCCACAGCTAGGCGCACTGCGCCCGTAAG

     AGCAGCAGCGTGCCGCCAAGACGCTCGAACAGGGCGACCATGCCCGCGACCAGCTGGTTG
5761 ---------+---------+---------+---------+---------+---------+ 5820
     TCGTCGTCGCACGGCGGTTCTGCGAGCTTGTCCCGCTGGTACGGGCGCTGGTCGACCAAC

     GTGCCGCCCTTGGCGAACCAGACGCCGCCGCGCCGTTCCAGCGCATGGATCAGCGCATAG
5821 ---------+---------+---------+---------+---------+---------+ 5880
     CACGGCGGGAACCGCTTGGTCTGCGGCGGCGCGGCAAGGTCGCGTACCTAGTCGCGTATC

     ATCGAGCTGGTCGAAAACGGGTTCCCGCCGACCAGCAGCGTGTGGAACGAGAAGGCCTGC
5881 ---------+---------+---------+---------+---------+---------+ 5940
     TAGCTCGACCAGCTTTTGCCCAAGGGCGGCTGGTCGTCGCACACCTTGCTCTTCCGGACG

     CGCAGATGCGGGTCCTGGATGAAGCGCGCCACCATGCTGTGGACCGAGCGGTATGCCTGC
5941 ---------+---------+---------+---------+---------+---------+ 6000
     GCGTCTACGCCCAGGACCTACTTCGCGCGGTGGTACGACACCTGGCTCGCCATACGGACG

     AGGCGCATCAGCGCCGGCGCGGCGTTCAGCATCTGGCCCAGCTTCAGGAAGGGCGTGGTC
6001 ---------+---------+---------+---------+---------+---------+ 6060
     TCCGCGTAGTCGCGGCCGCGCCGCAAGTCGTAGACCGGGTCGAAGTCCTTCCCGCACCAG
```

Fig. 24/7

```
     CCCAGCTTCAGATACCCCTCGCGATAGACCTCCTCGGCGTAATCGTGGAAGCGGCGATAG
6061 ---------+---------+---------+---------+---------+---------+ 6120
     GGGTCGAAGTCTATGGGGAGCGCTATCTGGAGGAGCCGCATTAGCACCTTCGCCGCTATC

     CCATCGACATCGGCGGGATTGAAGGAGCCGACCTGGCGGATCAGCTCGTCGTCGTCGTTC
6121 ---------+---------+---------+---------+---------+---------+ 6180
     GGTAGCTGTAGCCGCCCTAACTTCCTCGGCTGGACCGCCTAGTCGAGCAGCAGCAGCAAG

     ACGTATTCGAAGCTGCGGCCGTCCGCCCATGTCAGCCGGTAGAAGGGCGAGACCGGCAGC
6181 ---------+---------+---------+---------+---------+---------+ 6240
     TGCATAAGCTTCGACGCCGGCAGGCGGGTACAGTCGGCCATCTTCCCGCTCTGGCCGTCG

     AGCGTCACGTCACGCTCCATCGGTTGGCCGCTGAGGGCCCACAGCTCTCGCAGGCTGTCG
6241 ---------+---------+---------+---------+---------+---------+ 6300
     TCGCAGTGCAGTGCGAGGTAGCCAACCGGCGACTCCCGGGTGTCGAGAGCGTCCGACAGC

     GGGTCGGTCACGACCGTCGGGCCTGCATCGAAGACGTGGCCCTGATCGTTCCAGACATAG
6301 ---------+---------+---------+---------+---------+---------+ 6360
     CCCAGCCAGTGCTGGCAGCCCGGACGTAGCTTCTGCACCGGGACTAGCAAGGTCTGTATC

     GCGCGGCCGCCGGGCTTGTCGCGGGCCTCGACGATGGTGGTCGCGATGCCGGCCGATTGC
6361 ---------+---------+---------+---------+---------+---------+ 6420
     CGCGCCGGCGGCCCGAACAGCGCCCGGAGCTGCTACCACCAGCGCTACGGCCGGCTAACG

     AGGCGGATGGCAAGCGCAAGCCCGCCGAAACCTGCGCCGATGACGATGGCGGAACTCATG
6421 ---------+---------+---------+---------+---------+---------+ 6480
     TCCGCCTACCGTTCGCGTTCGGGCGGCTTTGGACGCGGCTACTGCTACCGCCTTGAGTAC

     CTCTCTCCTGCAGCAGGGGGCGTTCGGGCAGGCAGCGCACGGCCTGCGACAGCGGAATGG
6481 ---------+---------+---------+---------+---------+---------+ 6540
     GAGAGAGGACGTCGTCCCCCGCAAGCCCGTCCGTCGCGTGCCGGACGCTGTCGCCTTACC

     GCGGGCGTCCGGTGACGATGCGAAGCCGGTCGGCCAATGTCAGGCGCCCGGCATAGAAGC
6541 ---------+---------+---------+---------+---------+---------+ 6600
     CGCCCGCAGGCCACTGCTACGCTTCGGCCAGCCGGTTACAGTCCGCGGGCCGTATCTTCG

     GCTCGATCAGCGGCTGCGGCAGGCGGTAGAACCGCTGCAGCAGGCGATAGCGACGGTCGG
6601 ---------+---------+---------+---------+---------+---------+ 6660
     CGAGCTAGTCGCCGACGCCGTCCGCCATCTTGGCGACGTCGTCCGCTATCGCTGCCAGCC

     GCGGGCAGCCGCGGAACAGCATCCGGTTCAGCAGCCGCAGGAAGCGGTCGCGATCCGCGC
6661 ---------+---------+---------+---------+---------+---------+ 6720
     CGCCCGTCGGCGCCTTGTCGTAGGCCAAGTCGTCGGCGTCCTTCGCCAGCGCTAGGCGCG

     GATCGATGGCCCAGCCGCGCACCGCGCGACGGGCGGACGCGGTCGTCAGGTCGCGCGCCG
6721 ---------+---------+---------+---------+---------+---------+ 6780
     CTAGCTACCGGGTCGGCGCGTGGCGCGCTGCCCGCCTGCGCCAGCAGTCCAGCGCGCGGC

     CGATGGCATCCGCGACCTGCGCGGCATAGGGCAGCGAATATCCGGTGACGGGGTGGAACA
6781 ---------+---------+---------+---------+---------+---------+ 6840
     GCTACCGTAGGCGCTGGACGCGCCGTATCCCGTCGCTTATAGGCCACTGCCCCACCTTGT

     GCCCTGCCCCCAGCCCAACCGGCACCGCCCCCTGCGCGTGGTCGCGCCAGAAGCCTATGG
6841 ---------+---------+---------+---------+---------+---------+ 6900
     CGGGACGGGGGTCGGGTTGGCCGTGGCGGGGGACGCGCACCAGCGCGGTCTTCGGATACC

     CGTCATGGGCCAGCGCGATGGGCAGGATGCCCCTTTCGCGCCGCATCTCCTGCCCGGTCC
6901 ---------+---------+---------+---------+---------+---------+ 6960
     GCAGTACCCGGTCGCGCTACCCGTCCTACGGGGAAAGCGCGGCGTAGAGGACGGGCCAGG

     AGCCCCGCCTGGCGGCATAGTCCAGCGACGCCTGCGCCAGCGCGCCATCGTCCAGATCGC
6961 ---------+---------+---------+---------+---------+---------+ 7020
     TCGGGGCGGACCGCCGTATCAGGTCGCTGCGGACGCGGTCGCGCGGTAGCAGGTCTAGCG
```

Fig. 24/8

```
     CGCCGTCGCTGTAGCGCGTATCCTCGATCAGGATGCGGGTGGGACTGAAGGGCAGCAGAT
7021 ---------+---------+---------+---------+---------+---------+ 7080
     GCGGCAGCGACATCGCGCATAGGAGCTAGTCCTACGCCCACCCTGACTTCCCGTCGTCTA

     AGATGAAGCGGTACCCGTCCATCTGCGGAACGGTCGCGTCCATGATCATCGGGCGCTCGA
7081 ---------+---------+---------+---------+---------+---------+ 7140
     TCTACTTCGCCATGGGCAGGTAGACGCCTTGCCAGCGCAGGTACTAGTAGCCCGCGAGCT

     CGCCATGGGGGGCGTCGGTCTCGATCTCGACGCCCACGAATTTCTGGAAACCCACGGTCA
7141 ---------+---------+---------+---------+---------+---------+ 7200
     GCGGTACCCCCCGCAGCCAGAGCTAGAGCTGCGGGTGCTTAAAGACCTTTGGGTGCCAGT

     GGTGCGGGGTCTCGACGGCACCACGGGCGTCGATCACGCAGGCAGCCTCGATCCGCGAGC
7201 ---------+---------+---------+---------+---------+---------+ 7260
     CCACGCCCCAGAGCTGCCGTGGTGCCCGCAGCTAGTGCGTCCGTCGGAGCTAGGCGCTCG

     CGTCCGTCAGCGTCGCGCCGGTATCGTCCAGCGTCGCGACATGCGTATTCCACCGCAGAT
7261 ---------+---------+---------+---------+---------+---------+ 7320
     GCAGGCAGTCGCAGCGCGGCCATAGCAGGTCGCAGCGCTGTACGCATAAGGTGGCGTCTA

     CGACACCCTGCAGCAGCCCGATCAGCGCGCCCGCCTCGATCGAGCCATAGCCTGTCGTCA
7321 ---------+---------+---------+---------+---------+---------+ 7380
     GCTGTGGGACGTCGTCGGGCTAGTCGCGCGGGCGGAGCTAGCTCGGTATCGGACAGCAGT

     GGCGGCGCGAATGGTCGGGAAACGCGACCTCCTGATCCGTCCATTCGCCGCGACGAATGG
7381 ---------+---------+---------+---------+---------+---------+ 7440
     CCGCCGCGCTTACCAGCCCTTTGCGCTGGAGGACTAGGCAGGTAAGCGGCGCTGCTTACC

     GCGACAGGCGCGCCAGCCATTCGGGCGAAAGATCCGTGTCGTGGCAGGACCAGGTGTGCT
7441 ---------+---------+---------+---------+---------+---------+ 7500
     CGCTGTCCGCGCGGTCGGTAAGCCCGCTTTCTAGGCACAGCACCGTCCTGGTCCACACGA

     GGTCCGAGGGGCCGGACCGCGCGTCGAGCATCACGATGCGCGCATCCGGTCTGCGGTCGC
7501 ---------+---------+---------+---------+---------+---------+ 7560
     CCAGGCTCCCCGGCCTGGCGCGCAGCTCGTAGTGCTACGCGCGTAGGCCAGACGCCAGCG

     GAACGGCAAGCGCGATCAGCGCACCGGACAGCCCCGCGCCCGCGATCAGCAGATCATGGC
7561 ---------+---------+---------+---------+---------+---------+ 7620
     CTTGCCGTTCGCGCTAGTCGCGTGGCCTGTCGGGGCGCGGGCGCTAGTCGTCTAGTACCG

     TCATGTATTGCGATCCGCCCCTTCGCGGTCCTTCAGCAGCGCGCCCGAGCGTTTCAGCTC
7621 ---------+---------+---------+---------+---------+---------+ 7680
     AGTACATAACGCTAGGCGGGGAAGCGCCAGGAAGTCGTCGCGCGGGCTCGCAAAGTCGAG

     TGCCTTGAGGCTGTCGACCGAGGGCGCCCAGATGAAACCGAAGCTGACGCAGTTCTCGCG
7681 ---------+---------+---------+---------+---------+---------+ 7740
     ACGGAACTCCGACAGCTGGCTCCCGCGGGTCTACTTTGGCTTCGACTGCGTCAAGAGCGC

     GCCATGGACCGCGTGATGCATCCTGTGTGCCTGGTAGACGCGACGAAGATAGCCGCGCTT
7741 ---------+---------+---------+---------+---------+---------+ 7800
     CGGTACCTGGCGCACTACGTAGGACACACGGACCATCTGCGCTGCTTCTATCGGCGCGAA

     GGGGACATAGCGGAACGGCCAGCGCCCATGCACCAAGCCGTCATGCAGGAAATAGTAGAT
7801 ---------+---------+---------+---------+---------+---------+ 7860
     CCCCTGTATCGCCTTGCCGGTCGCGGGTACGTGGTTCGGCAGTACGTCCTTTATCATCTA

     CAGCCCGTAGCAGGTGACCCCCACCGCCAGCCACCAGGCCAGATCCGACCCCATCGCGCC
7861 ---------+---------+---------+---------+---------+---------+ 7920
     GTCGGGCATCGTCCACTGGGGGTGGCGGTCGGTGGTCCGGTCTAGGCTGGGGTAGCGCGG

     GATCGCGAACAGCACGATCGAGATTACCGCGAAGATGACGCCATAGAGGTCGTTCTTCTC
7921 ---------+---------+---------+---------+---------+---------+ 7980
     CTAGCGCTTGTCGTGCTAGCTCTAATGGCGCTTCTACTGCGGTATCTCCAGCAAGAAGAG
```

Fig. 24/9

```
     GAGCGCGTGGTCGTGATCCTCGTCGTGGTGCGATTTATGCCAGCCCCAGCCCAGGGGGCC
7981 ---------+---------+---------+---------+---------+---------+ 8040
     CTCGCGCACCAGCACTAGGAGCAGCACCACGCTAAATACGGTCGGGGTCGGGTCCCCCGG

     ATGCATGATCCACCGATGGACGGAGTAGGCCGTCAGCTCCATCGCGGCGACGGTCAGGAT
8041 ---------+---------+---------+---------+---------+---------+ 8100
     TACGTACTAGGTGGCTACCTGCCTCATCCGGCAGTCGAGGTAGCGCCGCTGCCAGTCCTA

     GACGGTCAGGATTGCGGCCCAAGTGCTCATGCCGGCCCCTTGCTTGATATGACAGGGAAC
8101 ---------+---------+---------+---------+---------+---------+ 8160
     CTGCCAGTCCTAACGCCGGGTTCACGAGTACGGCCGGGGAACGAACTATACTGTCCCTTG

     AGGCTACGCTGCCGCGCGGTGCATGACCAGCCCATCGGGGTGCGACCAAAGGGCATCGCG
8161 ---------+---------+---------+---------+---------+---------+ 8220
     TCCGATGCGACGGCGCGCCACGTACTGGTCGGGTAGCCCCACGCTGGTTTCCCGTAGCGC

     TGACATCTGCGTTCAGGGCTCATAGGCGGATCATCCGTGACATTCGCCGCCGAACGCGGC
8221 ---------+---------+---------+---------+---------+---------+ 8280
     ACTGTAGACGCAAGTCCCGAGTATCCGCCTAGTAGGCACTGTAAGCGGCGGCTTGCGCCG

     AGGCGCATCACGCGTTCCGTCGCTGGAAATATTAATGTTTTCCCGAAGATGGTCGGGGCG
8281 ---------+---------+---------+---------+---------+---------+ 8340
     TCCGCGTAGTGCGCAAGGCAGCGACCTTTATAATTACAAAAGGGCTTCTACCAGCCCCGC

     AGAGGATTCGAACCTCCGACCTACGGTACCCAAAACCGTCGCGCTACCAGGCTGCGCTAC
8341 ---------+---------+---------+---------+---------+---------+ 8400
     TCTCCTAAGCTTGGAGGCTGGATGCCATGGGTTTTGGCAGCGCGATGGTCCGACGCGATG

     GCCCCGACTGCGGAAGGCTTTAGCCGATTGTTCCGGCAAGGGAAAGACCTAGTCGCAGGC
8401 ---------+---------+---------+---------+---------+---------+ 8460
     CGGGGCTGACGCCTTCCGAAATCGGCTAACAAGGCCGTTCCCTTTCTGGATCAGCGTCCG

     CAGGACCGCATTGTCGCCCATGCCCGGATGCGCCATCGGCTGACCGGGCTTCAGGCCAAG
8461 ---------+---------+---------+---------+---------+---------+ 8520
     GTCCTGGCGTAACAGCGGGTACGGGCCTACGCGGTAGCCGACTGGCCCGAAGTCCGGTTC

     GCGATCCGCCTCTCCGCCCGCGATTTCGAGGACGAACAGCCGGTCGGGGTCCGGATCGCC
8521 ---------+---------+---------+---------+---------+---------+ 8580
     CGCTAGGCGGAGAGGCGGGCGCTAAAGCTCCTGCTTGTCGGCCAGCCCCAGGCCTAGCGG

     GACCGCCGCGCCCGGAATGGGCGTCTCGTCCAGCGGGCGCGCATTGCGGTGGATGTGGCG
8581 ---------+---------+---------+---------+---------+---------+ 8640
     CTGGCGGCGCGGGCCTTACCCGCAGAGCAGGTCGCCCGCGCGTAACGCCACCTACACCGC

     GATGACGCCGGTTTCATCCGCAAAGACCATGTCCAGCGGGATCAGTGTGTTGCGCATCCA
8641 ---------+---------+---------+---------+---------+---------+ 8700
     CTACTGCGGCCAAAGTAGGCGTTTCTGGTACAGGTCGCCCTAGTCACACAACGCGTAGGT

     GAAGGACACCGGCTGGGGCGATTCGTAGATGAACAGCATTCCGGTGCCCGCAGGCAGCTC
8701 ---------+---------+---------+---------+---------+---------+ 8760
     CTTCCTGTGGCCGACCCCGCTAAGCATCTACTTGTCGTAAGGCCACGGGCGTCCGTCGAG

     CTTGCGGAACATCAGGCCCTGCGCGCGCTCTTCGGGGCTGTCCGCGACCTCGACCCGAAA
8761 ---------+---------+---------+---------+---------+---------+ 8820
     GAACGCCTTGTAGTCCGGGACGCGCGCGAGAAGCCCCGACAGGCGCTGGAGCTGGGCTTT

     CCCGAGCGTTTCCGCACCGGTATCGACGACAAGACTGCCGGGCGCGCATTCCACCGCCGC
8821 ---------+---------+---------+---------+---------+---------+ 8880
     GGGCTCGCAAAGGCGTGGCCATAGCTGCTGTTCTGACGGCCCGCGCGTAAGGTGGCGGCG

     CGCGGCGGCGGGCATCAGGACCGCAAGAAGCGCTGCGGCCTTACTCGGCCACATGGGCAA
8881 ---------+---------+---------+---------+---------+---------+ 8940
     GCGCCGCCGCCCGTAGTCCTGGCGTTCTTCGCGACGCCGGAATGAGCCGGTGTACCCGTT

     GATAGGACTGCTCGGCGCCGAGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCG
8941 ---------+---------+---------+---------+---------+---------+ 9000
     CTATCCTGACGAGCCGCGGCTCTAGGGGGCCCGACGTCCTTAAGCTATAGTTCGAATAGC
```

Fig. 24/10

```
      ATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTA
9001 ---------+---------+---------+---------+---------+---------+ 9060
      TATGGCAGCTGGAGCTCCCCCCCGGGCCATGGGTCGAAAACAAGGGAAATCACTCCCAAT

      ATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
9061 ---------+---------+---------+---------+---------+---------+ 9120
      TAACGCGCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAG

      ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA
9121 ---------+---------+---------+---------+---------+---------+ 9180
      TGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACT

      GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
9181 ---------+---------+---------+---------+---------+---------+ 9240
      CACTCGATTGAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGAC

      TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
9241 ---------+---------+---------+---------+---------+---------+ 9300
      AGCACGGTCGACGTAATTACTTAGCCGGTTGCGCGCCCCTCTCCGCCAAACGCATAACCC

      CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
9301 ---------+---------+---------+---------+---------+---------+ 9360
      GCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGC

      GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGA
9361 ---------+---------+---------+---------+---------+---------+ 9420
      CATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCT

      AAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG
9421 ---------+---------+---------+---------+---------+---------+ 9480
      TTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGAC

      GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAG
9481 ---------+---------+---------+---------+---------+---------+ 9540
      CGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTC

      AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
9541 ---------+---------+---------+---------+---------+---------+ 9600
      TCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAG

      GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
9601 ---------+---------+---------+---------+---------+---------+ 9660
      CACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGC

      GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
9661 ---------+---------+---------+---------+---------+---------+ 9720
      CCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAA

      CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
9721 ---------+---------+---------+---------+---------+---------+ 9780
      GCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGG

      GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
9781 ---------+---------+---------+---------+---------+---------+ 9840
      CCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGG

      ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG
9841 ---------+---------+---------+---------+---------+---------+ 9900
      TGACCATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACC

      TGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
9901 ---------+---------+---------+---------+---------+---------+ 9960
      ACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGT
```

Fig. 24/11

```
      GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
9961  ---------+---------+---------+---------+---------+---------+ 10020
      CAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCG

      GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
10021 ---------+---------+---------+---------+---------+---------+ 10080
      CCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTA

      CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATT
10081 ---------+---------+---------+---------+---------+---------+ 10140
      GGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAA

      TTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
10141 ---------+---------+---------+---------+---------+---------+ 10200
      AACCAGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCA

      TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATC
10201 ---------+---------+---------+---------+---------+---------+ 10260
      AAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAG

      AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCC
10261 ---------+---------+---------+---------+---------+---------+ 10320
      TCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGG

      GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
10321 ---------+---------+---------+---------+---------+---------+ 10380
      CAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTAT

      CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGG
10381 ---------+---------+---------+---------+---------+---------+ 10440
      GGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCC

      GCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
10441 ---------+---------+---------+---------+---------+---------+ 10500
      CGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACG

      CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
10501 ---------+---------+---------+---------+---------+---------+ 10560
      GCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

      ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAA
10561 ---------+---------+---------+---------+---------+---------+ 10620
      TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTT

      CGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT
10621 ---------+---------+---------+---------+---------+---------+ 10680
      GCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCA

      CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA
10681 ---------+---------+---------+---------+---------+---------+ 10740
      GGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGT

      CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTAC
10741 ---------+---------+---------+---------+---------+---------+ 10800
      GACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATG

      TCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCA
10801 ---------+---------+---------+---------+---------+---------+ 10860
      AGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGT

      ATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGT
10861 ---------+---------+---------+---------+---------+---------+ 10920
      TATGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCA

      TCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
10921 ---------+---------+---------+---------+---------+---------+ 10980
      AGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGG
```

Fig. 24/12

```
      ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCA
10981 ---------+---------+---------+---------+---------+---------+ 11040
      TGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGT

      AAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
11041 ---------+---------+---------+---------+---------+---------+ 11100
      TTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTAT

      CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
11101 ---------+---------+---------+---------+---------+---------+ 11160
      GAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCG

      GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCC
11161 ---------+---------+---------+---------+---------+---------+ 11220
      CCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGG

      CGAAAAGTGCCAC
11221 ---------+--- 11233
      GCTTTTCACGGTG
```

Fig. 25

```
                                                  crtW1
     MetSerGlyArgLysProGlyThrThrGlyAspThrIleValAsnLeuGlyLeuThrAlaAlaIleLeuLeuCysTrpLeuValLeuHisAlaPheThrLeuTrpLeuLeuAspAlaAla
     ATGTCCGGTCGTAAACCGGGTACCACCGGTGACACCATCGTTAACCTGGGTCTGACCGCTGCTATCCTGCTGTGCTGGCTGGTTCTGCACGCTTTCACCCTGTGGCTGCTGGACGCTGCT
   1 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 120
     TACAGGCCAGCATTTGGCCCATGGTGGCCACTGTGGTAGCAATTGGACCCAGACTGGCGACGATAGGACGACACGACCGACCAAGACGTGCGAAAGTGGGACACCGACGACCTGCGACGA
                                                                                                           crtW2

                                                  crtW3
     AlaHisProLeuLeuAlaValLeuCysLeuAlaGlyLeuThrTrpLeuSerValGlyLeuPheIleIleAlaHisAspAlaMetHisGlySerValValProGlyArgProArgAlaAsn
     GCTCACCCGCTGCTGGCTGTTCTGTGCCTGGCTGGTCTGACCTGGCTGTCCGTTGGTCTGTTCATCATCGCTCACGACGCTATGCACGGTTCCGTTGTTCCGGGTCGTCCGCGGGCTAAC
 121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 240
     CGAGTGGGCGACGACCGACAAGACACGGACCGACCAGACTGGACCGACAGGCAACCAGACAAGTAGTAGCGAGTGCTGCGATACGTGCCAAGGCAACAAGGCCCAGCAGGCGCCCGATTG
                                                                                                           crtW4

                                                  crtW5
     AlaAlaIleGlyGlnLeuAlaLeuTrpLeuTyrAlaGlyPheSerTrpProLysLeuIleAlaLysHisMetThrHisHisArgHisAlaGlyThrAspAsnAspProAspPheGlyHis
     GCTGCTATCGGTCAGCTGGCTCTGTGGCTGTACGCTGGTTTCTCCTGGCCGAAACTGATCGCTAAACACATGACCCACCACCGTCACGCTGGTACCGACAACGACCCGGACTTCGGTCAC
 241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 360
     CGACGATAGCCAGTCGACCGAGACACCGACATGCGACCAAAGAGGACCGGCTTTGACTAGCGATTTGTGTACTGGGTGGTGGCAGTGCGACCATGGCTGTTGCTGGGCCTGAAGCCAGTG
                                                                                                           crtW6

                                                  crtW7
     GlyGlyProValArgTrpTyrGlySerPheValSerThrTyrPheGlyTrpArgGluGlyLeuLeuLeuProValIleValThrThrTyrAlaLeuIleLeuGlyAspArgTrpMetTyr
     GGTGGTCCGGTTCGTTGGTACGGTTCCTTCGTTTCCACCTACTTCGGTTGGCGTGAAGGTCTGCTGCTGCCGGTTATCGTTACCACCTACGCTCTGATCCTGGGTGACCGTTGGATGTAC
 361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 480
     CCACCAGGCCAAGCAACCATGCCAAGGAAGCAAAGGTGGATGAAGCCAACCGCACTTCCAGACGACGACGGCCAATAGCAATGGTGGATGCGAGACTAGGACCCACTGGCAACCTACATG
                                                                                                           crtW8

                                                  crtW9
     ValIlePheTrpProValProAlaValLeuAlaSerIleGlnIlePheValPheGlyThrTrpLeuProHisArgProGlyHisAspAspPheProAspArgHisAsnAlaArgSerThr
     GTTATCTTCTGGCCGGTTCCGGCTGTTCTGGCTTCCATCCAGATCTTCGTTTTCGGTACCTGGCTGCCGCACCGTCCGGGTCACGACGACTTCCCGGACCGTCACAACGCTCGTTCCACC
 481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 600
     CAATAGAAGACCGGCCAAGGCCGACAAGACCGAAGGTAGGTCTAGAAGCAAAAGCCATGGACCGACGGCGTGGCAGGCCCAGTGCTGCTGAAGGGCCTGGCAGTGTTGCGAGCAAGGTGG
                                                                                                           crtW10

                                                  crtW11
     GlyIleGlyAspProLeuSerLeuLeuThrCysPheHisPheGlyGlyTyrHisHisGluHisHisLeuHisProHisValProTrpTrpArgLeuProArgThrArgLysThrGlyGly
     GGTATCGGTGACCCGCTGTCCCTGCTGACCTGCTTCCACTTCGGTGGTTACCACCACGAACACCACCTGCACCCGCACGTTCCGTGGTGGCGTCTGCCGCGTACCCGTAAAACCGGTGGT
 601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 720
     CCATAGCCACTGGGCGACAGGGACGACTGGACGAAGGTGAAGCCACCAATGGTGGTGCTTGTGGTGGACGTGGGCGTGCAAGGCACCACCGCAGACGGCGCATGGGCATTTTGGCCACCA
                                                                                                           crtW12

     ArgAla
     CGTGCT
 721 ------ 726
     GCACGA
```

Fig. 26

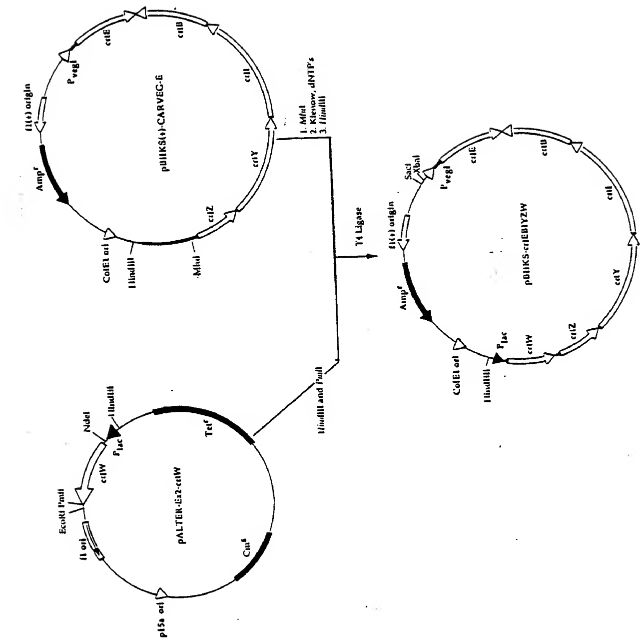

f1(+) origin
Amp^r
P_vegI
crtE
crtB
crtI
crtY
crtZ
ColE1 ori
HindIII
MluI
pBIIKS(+)-CARVEG-E
1. MluI
2. Klenow, dNTPs
3. HindIII
T4 Ligase
HindIII and PmlI
pBIIKS-crtEBIYZW
SacI
XbaI
P_lac
crtW
pALTER-Ex2-crtW
NdeI
Tet^r
Cm^R
f1 ori
EcoRI PmlI
p15a ori

Fig. 27

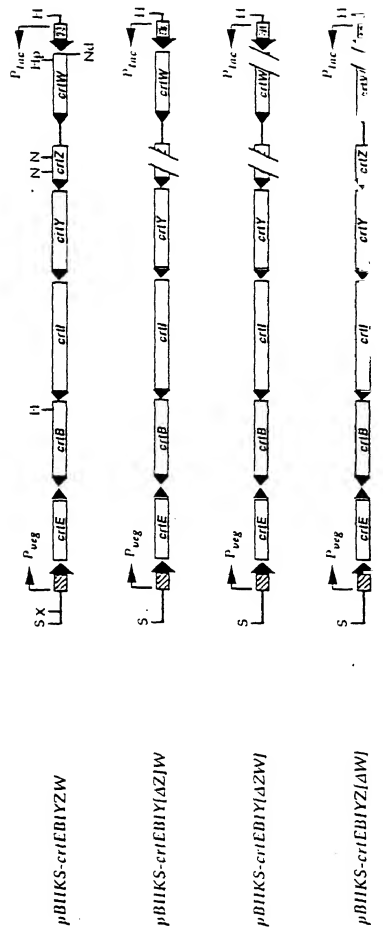

pBIIKS-crtEBIYZW
pBIIKS-crtEBIY[ΔZ]W
pBIIKS-crtEBIY[ΔZW]
pBIIKS-crtEBIYZ[ΔW]
S X
S
P_{veg}
P_{tac}
crtE
crtB
crtI
crtY
crtZ
crtW
H
Hp
Nd
N N

Fig. 28

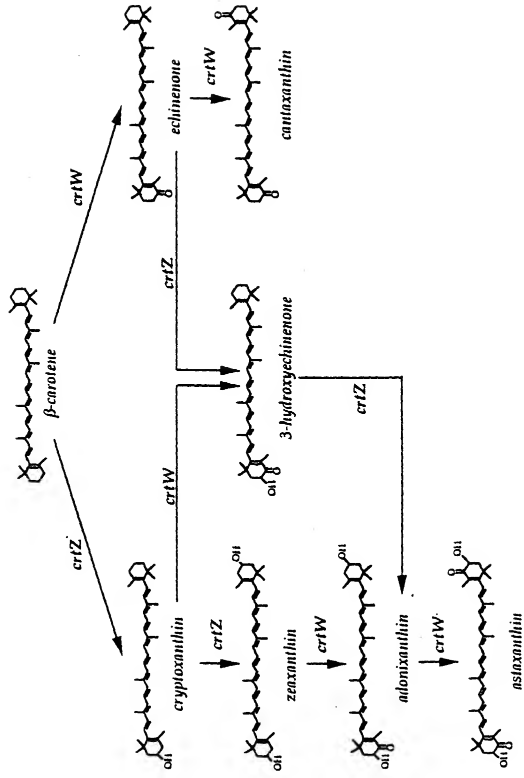

β-carotene
crtW
crtZ
echinenone
crtW
canthaxanthin
crtZ
crtW
3-hydroxyechinenone
crtZ
cryptoxanthin
crtZ
zeaxanthin
crtW
adonixanthin
crtW
astaxanthin

Fig. 29

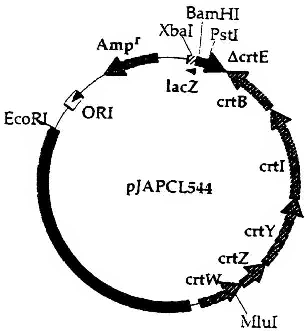

BamHI
XbaI
PstI
Amp^r
ΔcrtE
lacZ
crtB
EcoRI
ORI
crtI
pJAPCL544
crtY
crtZ
crtW
MluI

Fig. 30/1

```
     ACTGTAGTCTGCGCGGATCGCCGGTCCGGGGGACAAGATATGAGCGCACATGCCCTGCCC
  1  ---------+---------+---------+---------+---------+---------+  60
     TGACATCAGACGCGCCTAGCGGCCAGGCCCCCTGTTCTATACTCGCGTGTACGGGACGGG

     AAGGCAGATCTGACCGCCACCAGTTTGATCGTCTCGGGCGGCATCATCGCCGCGTGGCTG
 61  ---------+---------+---------+---------+---------+---------+  120
     TTCCGTCTAGACTGGCGGTGGTCAAACTAGCAGAGCCCGCCGTAGTAGCGGCGCACCGAC

     GCCCTGCATGTGCATGCGCTGTGGTTTCTGGACGCGGCGGCGCATCCCATCCTGGCGGTC
121  ---------+---------+---------+---------+---------+---------+  180
     CGGGACGTACACGTACGCGACACCAAAGACCTGCGCCGCCGCGTAGGGTAGGACCGCCAG

     GCGAATTTCCTGGGGCTGACCTGGCTGTCGGTCGGTCTGTTCATCATCGCGCATGACGCG
181  ---------+---------+---------+---------+---------+---------+  240
     CGCTTAAAGGACCCCGACTGGACCGACAGCCAGCCAGACAAGTAGTAGCGCGTACTGCGC

     ATGCATGGGTCGGTCGTGCCGGGGCGCCCGCGCGCCAATGCGGCGATGGGCCAGCTTGTC
241  ---------+---------+---------+---------+---------+---------+  300
     TACGTACCCAGCCAGCACGGCCCCGCGGGCGCGCGGTTACGCCGCTACCCGGTCGAACAG

     CTGTGGCTGTATGCCGGATTTTCCTGGCGCAAGATGATCGTCAAGCACATGGCCCATCAT
301  ---------+---------+---------+---------+---------+---------+  360
     GACACCGACATACGGCCTAAAAGGACCGCGTTCTACTAGCAGTTCGTGTACCGGGTAGTA

     CGCCATGCCGGAACCGACGACGACCCAGATTTCGACCATGGCGGCCCGGTCCGCTGGTAC
361  ---------+---------+---------+---------+---------+---------+  420
     GCGGTACGGCCTTGGCTGCTGCTGGGTCTAAAGCTGGTACCGCCGGGCCAGGCGACCATG

     GCCCGCTTCATCGGCACCTATTTCGGCTGGCGCGAGGGGCTGCTGCTGCCCGTCATCGTG
421  ---------+---------+---------+---------+---------+---------+  480
     CGGGCGAAGTAGCCGTGGATAAAGCCGACCGCGCTCCCCGACGACGACGGGCAGTAGCAC

     ACGGTCTATGCGCTGATGTTGGGGGATCGCTGGATGTACGTGGTCTTCTGGCCGTTGCCG
481  ---------+---------+---------+---------+---------+---------+  540
     TGCCAGATACGCGACTACAACCCCCTAGCGACCTACATGCACCAGAAGACCGGCAACGGC

     TCGATCCTGGCGTCGATCCAGCTGTTCGTGTTCGGCATCTGGCTGCCGCACCGCCCCGGC
541  ---------+---------+---------+---------+---------+---------+  600
     AGCTAGGACCGCAGCTAGGTCGACAAGCACAAGCCGTAGACCGACGGCGTGGCGGGGCCG

     CACGACGCGTTCCCGGACCGCCACAATGCGCGGTCGTCGCGGATCAGCGACCCCGTGTCG
601  ---------+---------+---------+---------+---------+---------+  660
     GTGCTGCGCAAGGGCCTGGCGGTGTTACGCGCCAGCAGCGCCTAGTCGCTGGGGCACAGC

     CTGCTGACCTGCTTTCACTTTGGCGGTTATCATCACGAACACCACCTGCACCCGACGGTG
661  ---------+---------+---------+---------+---------+---------+  720
     GACGACTGGACGAAAGTGAAACCGCCAATAGTAGTGCTTGTGGTGGACGTGGGCTGCCAC

     CCTTGGTGGCGCCTGCCCAGCACCCGCACCAAGGGGGACACCGCATGACCAATTTCCTGA
721  ---------+---------+---------+---------+---------+---------+  780
     GGAACCACCGCGGACGGGTCGTGGGCGTGGTTCCCCCTGTGGCGTACTGGTTAAAGGACT

     TCGTCGTCGCCACCGTGCTGGTGATGGAGCTGACGGCCTATTCCGTCCACCGCTGGATCA
781  ---------+---------+---------+---------+---------+---------+  840
     AGCAGCAGCGGTGGCACGACCACTACCTCGACTGCCGGATAAGGCAGGTGGCGACCTAGT
```

Fig. 30/2

```
     TGCACGGCCCCTTGGGCTGGGGCTGGCACAAGTCCCACCACGAGGAACACGACCACGCGC
841  ---------+---------+---------+---------+---------+---------+  900
     ACGTGCCGGGGAACCCGACCCCGACCGTGTTCAGGGTGGTGCTCCTTGTGCTGGTGCGCG

     TGGAAAAGAACGACCTGTACGGCCTGGTCTTTGCGGTGATCGCCACGGTGCTGTTCACGG
901  ---------+---------+---------+---------+---------+---------+  960
     ACCTTTTCTTGCTGGACATGCCGGACCAGAAACGCCACTAGCGGTGCCACGACAAGTGCC

     TGGGCTGGATCTGGGCACCGGTCCTGTGGTGGATCGCCTTGGGCATGACCGTCTACGGGC
961  ---------+---------+---------+---------+---------+---------+  1020
     ACCCGACCTAGACCCGTGGCCAGGACACCACCTAGCGGAACCCGTACTGGCAGATGCCCG

     TGATCTATTTCGTCCTGCATGACGGGCTGGTGCATCAGCGCTGGCCGTTCCGCTATATCC
1021 ---------+---------+---------+---------+---------+---------+  1080
     ACTAGATAAAGCAGGACGTACTGCCCGACCACGTAGTCGCGACCGGCAAGGCGATATAGG

     CTCGCAAGGGCTATGCCAGACGCCTGTATCAGGCCCACCGCCTGCACCACGCGGTCGAGG
1081 ---------+---------+---------+---------+---------+---------+  1140
     GAGCGTTCCCGATACGGTCTGCGGACATAGTCCGGGTGGCGGACGTGGTGCGCCAGCTCC

     GGCGCGACCATTGCGTCAGCTTCGGCTTCATCTATGCGCCGCCGGTCGACAAGCTGAAGC
1141 ---------+---------+---------+---------+---------+---------+  1200
     CCGCGCTGGTAACGCAGTCGAAGCCGAAGTAGATACGCGGCGGCCAGCTGTTCGACTTCG

     AGGACCTGAAGACGTCGGGCGTGCTGCGGGCCGAGGCGCAGGAGCGCACGTGACCCATGA
1201 ---------+---------+---------+---------+---------+---------+  1260
     TCCTGGACTTCTGCAGCCCGCACGACGCCCGGCTCCGCGTCCTCGCGTGCACTGGGTACT

     C
1261 -  1261
     G
```

Fig. 31

```
      ATGAGCGCACATGCCCTGCCCAAGGCAGATCTGACCGCCACCAGTTTGATCGTCTCGGGC
  1   ---------+---------+---------+---------+---------+---------+  60
      TACTCGCGTGTACGGGACGGGTTCCGTCTAGACTGGCGGTGGTCAAACTAGCAGAGCCCG

      GGCATCATCGCCGCGTGGCTGGCCCTGCATGTGCATGCGCTGTGGTTTCTGGACGCGGCG
 61   ---------+---------+---------+---------+---------+---------+  120
      CCGTAGTAGCGGCGCACCGACCGGGACGTACACGTACGCGACACCAAAGACCTGCGCCGC

      GCGCATCCCATCCTGGCGGTCGCGAATTTCCTGGGGCTGACCTGGCTGTCGGTCGGTCTG
121   ---------+---------+---------+---------+---------+---------+  180
      CGCGTAGGGTAGGACCGCCAGCGCTTAAAGGACCCCGACTGGACCGACAGCCAGCCAGAC

      TTCATCATCGCGCATGACGCGATGCATGGGTCGGTCGTGCCGGGGCGCCCGCGCGCCAAT
181   ---------+---------+---------+---------+---------+---------+  240
      AAGTAGTAGCGCGTACTGCGCTACGTACCCAGCCAGCACGGCCCCGCGGGCGCGCGGTTA

      GCGGCGATGGGCCAGCTTGTCCTGTGGCTGTATGCCGGATTTTCCTGGCGCAAGATGATC
241   ---------+---------+---------+---------+---------+---------+  300
      CGCCGCTACCCGGTCGAACAGGACACCGACATACGGCCTAAAAGGACCGCGTTCTACTAG

      GTCAAGCACATGGCCCATCATCGCCATGCCGGAACCGACGACGACCCAGATTTCGACCAT
301   ---------+---------+---------+---------+---------+---------+  360
      CAGTTCGTGTACCGGGTAGTAGCGGTACGGCCTTGGCTGCTGCTGGGTCTAAAGCTGGTA

      GGCGGCCCGGTCCGCTGGTACGCCCGCTTCATCGGCACCTATTTCGGCTGGCGCGAGGGG
361   ---------+---------+---------+---------+---------+---------+  420
      CCGCCGGGCCAGGCGACCATGCGGGCGAAGTAGCCGTGGATAAAGCCGACCGCGCTCCCC

      CTGCTGCTGCCCGTCATCGTGACGGTCTATGCGCTGATGTTGGGGGATCGCTGGATGTAC
421   ---------+---------+---------+---------+---------+---------+  480
      GACGACGACGGGCAGTAGCACTGCCAGATACGCGACTACAACCCCCTAGCGACCTACATG

      GTGGTCTTCTGGCCGTTGCCGTCGATCCTGGCGTCGATCCAGCTGTTCGTGTTCGGCATC
481   ---------+---------+---------+---------+---------+---------+  540
      CACCAGAAGACCGGCAACGGCAGCTAGGACCGCAGCTAGGTCGACAAGCACAAGCCGTAG

      TGGCTGCCGCACCGCCCCGGCCACGACGCGTTCCCGGACCGCCACAATGCGCGGTCGTCG
541   ---------+---------+---------+---------+---------+---------+  600
      ACCGACGGCGTGGCGGGGCCGGTGCTGCGCAAGGGCCTGGCGGTGTTACGCGCCAGCAGC

      CGGATCAGCGACCCCGTGTCGCTGCTGACCTGCTTTCACTTTGGCGGTTATCATCACGAA
601   ---------+---------+---------+---------+---------+---------+  660
      GCCTAGTCGCTGGGGCACAGCGACGACTGGACGAAAGTGAAACCGCCAATAGTAGTGCTT

      CACCACCTGCACCCGACGGTGCCTTGGTGGCGCCTGCCCAGCACCCGCACCAAGGGGGAC
661   ---------+---------+---------+---------+---------+---------+  720
      GTGGTGGACGTGGGCTGCCACGGAACCACCGCGGACGGGTCGTGGGCGTGGTTCCCCCTG

      ACCGCATGA
721   ---------  729
      TGGCGTACT
```

Fig. 32

```
  1  MSAHALPKAD LTATSLIVSG GIIAAWLALH VHALWFLDAA AHPILAVANF
 51  LGLTWLSVGL FIIAHDAMHG SVVPGRPRAN AAMGQLVLWL YAGFSWRKMI
101  VKHMAHHRHA GTDDDPDFDH GGPVRWYARF IGTYFGWREG LLLPVIVTVY
151  ALMLGDRWMY VVFWPLPSIL ASIQLFVFGI WLPHRPGHDA FPDRHNARSS
201  RISDPVSLLT CFHFGGYHHE HHLHPTVPWW RLPSTRTKGD TA*
```

Fig. 33

```
    ATGACCAATTTCCTGATCGTCGTCGCCACCGTGCTGGTGATGGAGCTGACGGCCTATTCC
1   ---------+---------+---------+---------+---------+---------+  60
    TACTGGTTAAAGGACTAGCAGCAGCGGTGGCACGACCACTACCTCGACTGCCGGATAAGG

    GTCCACCGCTGGATCATGCACGGCCCCTTGGGCTGGGGCTGGCACAAGTCCCACCACGAG
61  ---------+---------+---------+---------+---------+---------+  120
    CAGGTGGCGACCTAGTACGTGCCGGGGAACCCGACCCCGACCGTGTTCAGGGTGGTGCTC

    GAACACGACCACGCGCTGGAAAAGAACGACCTGTACGGCCTGGTCTTTGCGGTGATCGCC
121 ---------+---------+---------+---------+---------+---------+  180
    CTTGTGCTGGTGCGCGACCTTTTCTTGCTGGACATGCCGGACCAGAAACGCCACTAGCGG

    ACGGTGCTGTTCACGGTGGGCTGGATCTGGGCACCGGTCCTGTGGTGGATCGCCTTGGGC
181 ---------+---------+---------+---------+---------+---------+  240
    TGCCACGACAAGTGCCACCCGACCTAGACCCGTGGCCAGGACACCACCTAGCGGAACCCG

    ATGACCGTCTACGGGCTGATCTATTTCGTCCTGCATGACGGGCTGGTGCATCAGCGCTGG
241 ---------+---------+---------+---------+---------+---------+  300
    TACTGGCAGATGCCCGACTAGATAAAGCAGGACGTACTGCCCGACCACGTAGTCGCGACC

    CCGTTCCGCTATATCCCTCGCAAGGGCTATGCCAGACGCCTGTATCAGGCCCACCGCCTG
301 ---------+---------+---------+---------+---------+---------+  360
    GGCAAGGCGATATAGGGAGCGTTCCCGATACGGTCTGCGGACATAGTCCGGGTGGCGGAC

    CACCACGCGGTCGAGGGGCGCGACCATTGCGTCAGCTTCGGCTTCATCTATGCGCCGCCG
361 ---------+---------+---------+---------+---------+---------+  420
    GTGGTGCGCCAGCTCCCCGCGCTGGTAACGCAGTCGAAGCCGAAGTAGATACGCGGCGGC

    GTCGACAAGCTGAAGCAGGACCTGAAGACGTCGGGCGTGCTGCGGGCCGAGGCGCAGGAG
421 ---------+---------+---------+---------+---------+---------+  480
    CAGCTGTTCGACTTCGTCCTGGACTTCTGCAGCCCGCACGACGCCCGGCTCCGCGTCCTC

    CGCACG
481 ------  486
    GCGTGC
```

Fig. 34

```
  1  MTNFLIVVAT VLVMELTAYS VHRWIMHGPL GWGWHKSHHE EHDHALEKND
 51  LYGLVFAVIA TVLFTVGWIW APVLWWIALG MTVYGLIYFV LHDGLVHQRW
101  PFRYIPRKGY ARRLYQAHRL HHAVEGRDHC VSFGFIYAPP VDKLKQDLKT
151  SGVLRAEAQE RT
```

Fig. 35

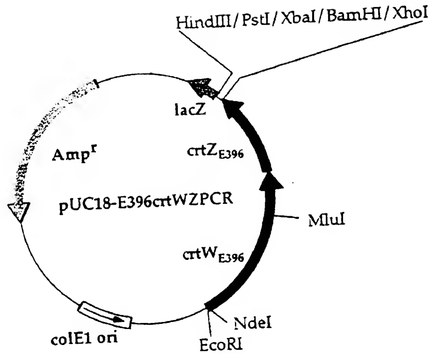
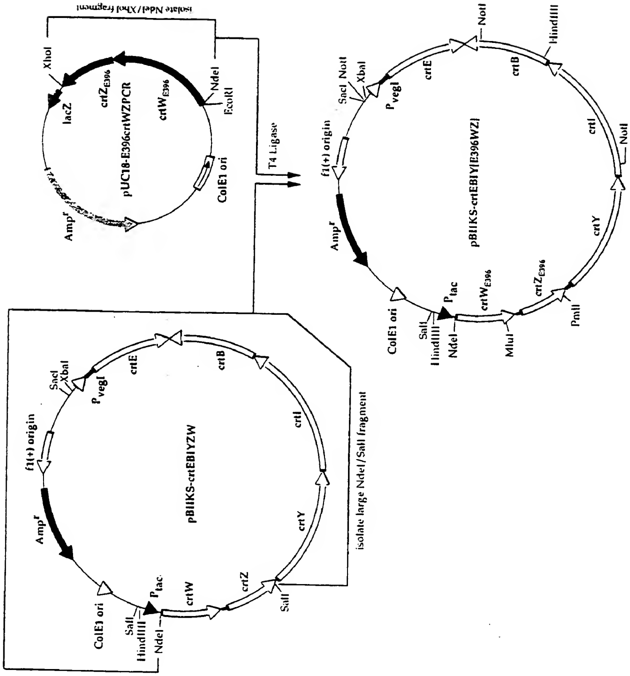

HindIII/PstI/XbaI/BamHI/XhoI
lacZ
Amp^r
crtZ E396
pUC18-E396crtWZPCR
MluI
crtW E396
NdeI
colE1 ori
EcoRI

Fig. 36



pUC18-E396crtWZPCR
isolate NdeI/XhoI fragment
XhoI
lacZ
crtZ E396
crtW E396
NdeI
EcoRI
ColE1 ori
Amp r
T4 Ligase
pBIIKS-crtEBIY[E396WZ]
NotI
HindIII
crtB
crtE
SacI NotI
XbaI
P vegI
f1(+) origin
crtI
NotI
crtY
crtZ E396
crtW E396
PmlI
MluI
P tac
ColE1 ori
SalI
HindIII
NdeI
pBIIKS-crtEBIYZW
isolate large NdeI/SalI fragment
SacI
XbaI
crtW
crtZ
SalI

Fig. 37

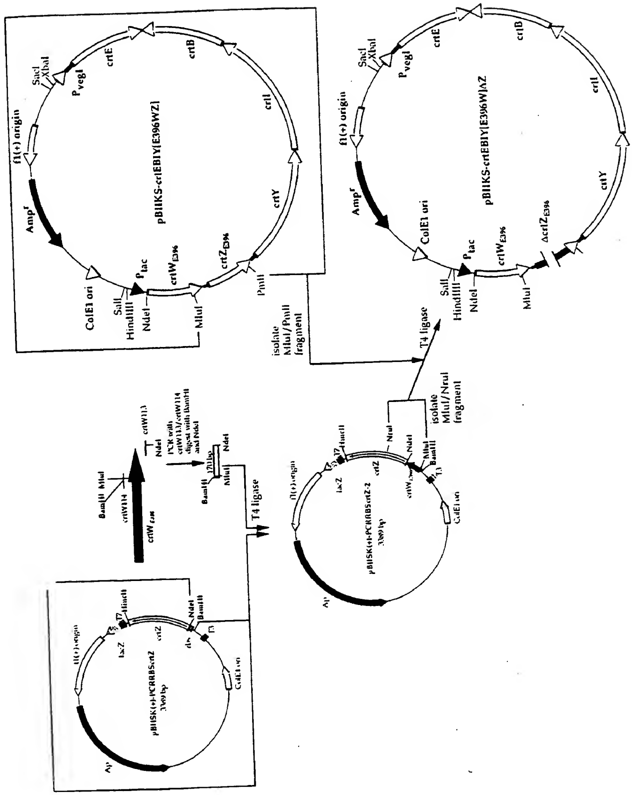

pBIIKS-crtEBIY[E396WZ]
f1(+) origin
SacI
XbaI
P vegI
crtE
crtB
crtI
crtY
crtZ E396
crtW E396
P tac
ColE1 ori
Amp r
SalI
HindIII
NdeI
MluI
PmlI
isolate MluI/PmlI fragment
T4 ligase
pBIIKS-crtEBIY[E396W]ΔZ
ΔcrtZ E396
isolate MluI/NruI fragment
NruI
BamHI
T4 ligase

Fig. 38/1

```
       CTGCAGGTCTGACACGGCCAGAAGGCCGCGCCGCGGGcCGGGGGCCGCcGCATCGCGACC
   1   ---------+---------+---------+---------+---------+---------+   60
       GACGTCCAGACTGTGCCGGTCTTCCGGCGCGGCGCCCgGCCCCCGGCGgCGTAGCGCTGG

       GGTATCCTTGCCAAGCGCCGCCTGGTCGCCCACaACGTCCAGCAGGTCGTCATAGGACTG
  61   ---------+---------+---------+---------+---------+---------+  120
       CCATAGGAACGGTTCGCGGCGGACCAGCGGGTGtTGCAGGTCGTCCAGCAGTATCCTGAC

       GAACACCCGGCCCAGCTGACGGCCAAAGTCGATCATCTGaGTCTGCTCCTCGGCGTCGAA
 121   ---------+---------+---------+---------+---------+---------+  180
       CTTGTGGGCCGGGTCGACTGCCGGTTTCAGCTAGTAGACtCAGACGAGGAGCCGCAGCTT

       CTCCTTGATCACGGCCAGCATCTCCAGCCCGGCGATGAACAGCACGCCGGTCTTCAGGTC
 181   ---------+---------+---------+---------+---------+---------+  240
       GAGGAACTAGTGCCGGTCGTAGAGGTCGGGCCGCTACTTGTCGTGCGGCCAGAAGTCCAG

       CTGTTCCTGTTCGACCCCCGCGCCGTTCTTGGCCGCGTGCAGGTCCAGGTCCTGGCCGGC
 241   ---------+---------+---------+---------+---------+---------+  300
       GACAAGGACAAGCTGGGGGCGCGGCAAGAACCGGCGCACGTCCAGGTCCAGGACCGGCCG

       GCACAGGCCCTGCGGCCCCAGGGACCGCGACAGGATCCgcaccagctgcgcccgcaccgt
 301   ---------+---------+---------+---------+---------+---------+  360
       CGTGTCCGGGACGCCGGGGTCCCTGGCGCTGTCCTAGGcgtggtcgacgcgggcgtggca

       gcccgacgcgccgcgcgcaccggccagcagggccatcgcctcggtgatcagggcgatgcc
 361   ---------+---------+---------+---------+---------+---------+  420
       cgggctgcgcggcgcgcgtggccggtcgtcccggtagcggagccactagtcccgctacgg

       gcctagcacggcgcggctttcgccatgcgccacatgggtcgcgggctggccgcggcgcag
 421   ---------+---------+---------+---------+---------+---------+  480
       cggatcgtgccgcgccgaaagcggtacgcggtgtacccagcgcccgaccggcgccgcgtc

       cccggcatcgtccatgcagggcaggtcgtcgaagatcagcgatgcggcatgcaccatctc
 481   ---------+---------+---------+---------+---------+---------+  540
       gggccgtagcaggtacgtcccgtccagcagcttctagtcgctacgccgtacgtggtagag

       gaccgcgcaggcggcgtcgacgatcgtgtcgcagaccccgcccgaggcttctgccgcaag
 541   ---------+---------+---------+---------+---------+---------+  600
       ctggcgcgtccgccgcagctgctagcacagcgtctggggcgggctccgaagacggcgttc

       cagcatcagcatgccgcggaaacgcttgcccgacgacagcgcgccatggctcatggccgg
 601   ---------+---------+---------+---------+---------+---------+  660
       gtcgtagtcgtacggcgcctttgcgaacgggctgctgtcgcgcggtaccgagtaccggcc

       gccgagcggctgcgacacggcaccgaatccctgggcgatctcctcaagtctggtctgcag
 661   ---------+---------+---------+---------+---------+---------+  720
       cggctcgccgacgctgtgccgtggcttagggacccgctagaggagttcagaccagacgtc

       aagggtggcgtggatcgggttgacgtctcgtctcatcagtgccttcgcgcttgggttctg
 721   ---------+---------+---------+---------+---------+---------+  780
       ttcccaccgcacctagcccaactgcagagcagagtagtcacggaagcgcgaacccaagac

       accaggcgggaaggtcaggccggggcggcaccccgtgacccgtcatccaccgtcaacagt
 781   ---------+---------+---------+---------+---------+---------+  840
       tggtccgcccttccagtccggccccgccgtggggcactgggcagtaggtggcagttgtca

       ccccatgttggaaggcttcacgcccgattgcgagcctttcgacggcgacgcggggtcgc
 841   ---------+---------+---------+---------+---------+---------+  900
       ggggtacaaccttccgaagtgcgggctaacgctcggaaaagctgccgctgcgccccagcg

       gcggcaatttntccaacaaggtcagtggaccggcgcgccgatggccgcgcgcagccaggc
 901   ---------+---------+---------+---------+---------+---------+  960
       cgccgttaaanaggttgttccagtcacctggccgcgcggctaccggcgcgcgtcggtccg

       atccttggccggaaacacccgcgccgcatcatgatcggccaggatcgtccggcgcgcggc
 961   ---------+---------+---------+---------+---------+---------+ 1020
```

Fig. 38/2

```
     taggaaccggcctttgtgggcgcggcgtagtactagccggtcctagcaggccgcgcgccg

     gcggcgcaggtcggccgcgtcacccggattgtcaagcacccaggccatcgcgtccgcgac
1021 ---------+---------+---------+---------+---------+---------+ 1080
     cgccgcgtccagccggcgcagtgggcctaacagttcgtgggtccggtagcgcaggcgctg

     ctcgtccgcgtcgtccatgtcgacgatcaggccgttctccatgtcgcggaccagttcgcg
1081 ---------+---------+---------+---------+---------+---------+ 1140
     gagcaggcgcagcaggtacagctgctagtccggcaagaggtacagcgcctggtcaagcgc

     caccggggcggtgttcgatcgatcaccaggcatccggtggccatcgcctcggacagggac
1141 ---------+---------+---------+---------+---------+---------+ 1200
     gtggccccgccacaagctagctagtggtccgtaggccaccggtagcggagcctgtccctg

     caggaggtgacgaagggctcggtgaaatagacatgcgcgtgcgaggcctgcag
1201 ---------+---------+---------+---------+---------+--- 1253
     gtcctccactgcttcccgagccactttatctgtacgcgcacgctccggacgtc
```

Fig. 39

```
     ATGAGACGAGACGTCAACCCGATCCACGCCACCCTTCTGCAGACCAGACTTGAGGAGATC
  1  ---------+---------+---------+---------+---------+---------+  60
     TACTCTGCTCTGCAGTTGGGCTAGGTGCGGTGGGAAGACGTCTGGTCTGAACTCCTCTAG

     GCCCAGGGATTCGGTGCCGTGTCGCAGCCGCTCGGCCCGGCCATGAGCCATGGCGCGCTG
 61  ---------+---------+---------+---------+---------+---------+  120
     CGGGTCCCTAAGCCACGGCACAGCGTCGGCGAGCCGGGCCGGTACTCGGTACCGCGCGAC

     TCGTCGGGCAAGCGTTTCCGCGGCATGCTGATGCTGCTTGCGGCAGAAGCCTCGGGCGGG
121  ---------+---------+---------+---------+---------+---------+  180
     AGCAGCCCGTTCGCAAAGGCGCCGTACGACTACGACGAACGCCGTCTTCGGAGCCCGCCC

     GTCTGCGACACGATCGTCGACGCCGCCTGCGCGGTCGAGATGGTGCATGCCGCATCGCTG
181  ---------+---------+---------+---------+---------+---------+  240
     CAGACGCTGTGCTAGCAGCTGCGGCGGACGCGCCAGCTCTACCACGTACGGCGTAGCGAC

     ATCTTCGACGACCTGCCCTGCATGGACGATGCCGGGCTGCGCCGCGGCCAGCCCGCGACC
241  ---------+---------+---------+---------+---------+---------+  300
     TAGAAGCTGCTGGACGGGACGTACCTGCTACGGCCCGACGCGGCGCCGGTCGGGCGCTGG

     CATGTGGCGCATGGCGAAAGCCGCGCCGTGCTAGGCGGCATCGCCCTGATCACCGAGGCG
301  ---------+---------+---------+---------+---------+---------+  360
     GTACACCGCGTACCGCTTTCGGCGCGGCACGATCCGCCGTAGCGGGACTAGTGGCTCCGC

     ATGGCCCTGCTGGCCGGTGCGCGCGGCGCGTCGGGCACGGTGCGGGCGCAGCTGGTGCGG
361  ---------+---------+---------+---------+---------+---------+  420
     TACCGGGACGACCGGCCACGCGCGCCGCGCAGCCCGTGCCACGCCCGCGTCGACCACGCC

     ATCCTGTCGCGGTCCCTGGGGCCGCAGGGCCTGTGCGCCGGCCAGGACCTGGACCTGCAC
421  ---------+---------+---------+---------+---------+---------+  480
     TAGGACAGCGCCAGGGACCCCGGCGTCCCGGACACGCGGCCGGTCCTGGACCTGGACGTG

     GCGGCCAAGAACGGCGCGGGGGTCGAACAGGAACAGGACCTGAAGACCGGCGTGCTGTTC
481  ---------+---------+---------+---------+---------+---------+  540
     CGCCGGTTCTTGCCGCGCCCCCAGCTTGTCCTTGTCCTGGACTTCTGGCCGCACGACAAG

     ATCGCCGGGCTGGAGATGCTGGCCGTGATCAAGGAGTTCGACGCCGAGGAGCAGACTCAG
541  ---------+---------+---------+---------+---------+---------+  600
     TAGCGGCCCGACCTCTACGACCGGCACTAGTTCCTCAAGCTGCGGCTCCTCGTCTGAGTC

     ATGATCGACTTTGGCCGTCAGCTGGGCCGGGTGTTCCAGTCCTATGACGACCTGCTGGAC
601  ---------+---------+---------+---------+---------+---------+  660
     TACTAGCTGAAACCGGCAGTCGACCCGGCCCACAAGGTCAGGATACTGCTGGACGACCTG

     GTTGTGGGCGACCAGGCGGCGCTTGGCAAGGATACCGGTCGCGATGCGGCGGCCCCCGGC
661  ---------+---------+---------+---------+---------+---------+  720
     CAACACCCGCTGGTCCGCCGCGAACCGTTCCTATGGCCAGCGCTACGCCGCCGGGGGCCG

     CCGCGGCGCGGCCTTCTGGCCGTGTCAGACCTGCAGAACGTGTCCCGTCACTATGAGGCC
721  ---------+---------+---------+---------+---------+---------+  780
     GGCGCCGCGCCGGAAGACCGGCACAGTCTGGACGTCTTGCACAGGGCAGTGATACTCCGG

     AGCCGCGCCCAGCTGGACGCGATGCTGCGCAGCAAGCGCCTTCAGGCTCCGGAAATCGCG
781  ---------+---------+---------+---------+---------+---------+  840
     TCGGCGCGGGTCGACCTGCGCTACGACGCGTCGTTCGCGGAAGTCCGAGGCCTTTAGCGC

     GCCCTGCTGGAACGGGTTCTGCCCTACGCCGCGCGCGCCTAG
841  ---------+---------+---------+---------+--  882
     CGGGACGACCTTGCCCAAGACGGGATGCGGCGCGCGCGGATC
```

Fig. 40

1 MRRDVNPIHA TLLQTRLEEI AQGFGAVSQP LGPAMSHGAL SSGKRFRGML

51 MLLAAEASGG VCDTIVDAAC AVEMVHAASL IFDDLPCMDD AGLRRGQPAT

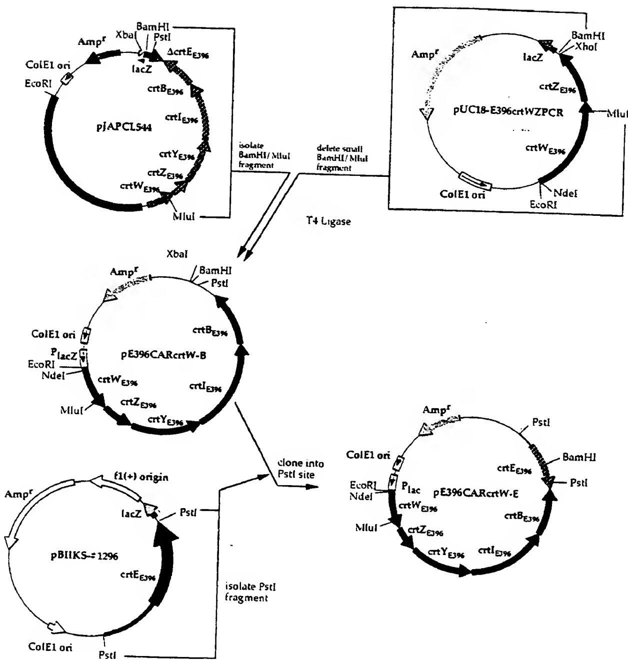
101 HVAHGESRAV LGGIALITEA MALLAGARGA SGTVRAQLVR ILSRSLGPQG

151 LCAGQDLDLH AAKNGAGVEQ EQDLKTGVLF IAGLEMLAVI KEFDAEEQTQ

201 MIDFGRQLGR VFQSYDDLLD VVGDQAALGK DTGRDAAAPG PRRGLLAVSD

251 LQNVSRHYEA SRAQLDAMLR SKRLQAPEIA ALLERVLPYA ARA*

Fig. 41



pJAPCL544
Amp^r
XbaI
BamHI
PstI
ΔcrtE E396
lacZ
ColE1 ori
EcoRI
crtB E396
crtI E396
crtY E396
crtZ E396
crtW E396
MluI
isolate BamHI/MluI fragment
delete small BamHI/MluI fragment
pUC18-E396crtWZPCR
Amp^r
lacZ
BamHI
XhoI
crtZ E396
MluI
crtW E396
ColE1 ori
NdeI
EcoRI
T4 Ligase
pE396CARcrtW-B
XbaI
Amp^r
BamHI
PstI
crtB E396
ColE1 ori
P lacZ
EcoRI
NdeI
crtW E396
crtI E396
crtZ E396
MluI
crtY E396
clone into PstI site
pBIIKS-1296
f1(+) origin
Amp^r
lacZ
PstI
crtE E396
ColE1 ori
PstI
isolate PstI fragment
pE396CARcrtW-E
Amp^r
PstI
ColE1 ori
BamHI
crtE E396
EcoRI
NdeI
P lac
PstI
crtW E396
crtB E396
MluI
crtZ E396
crtY E396
crtI E396

Fig. 42

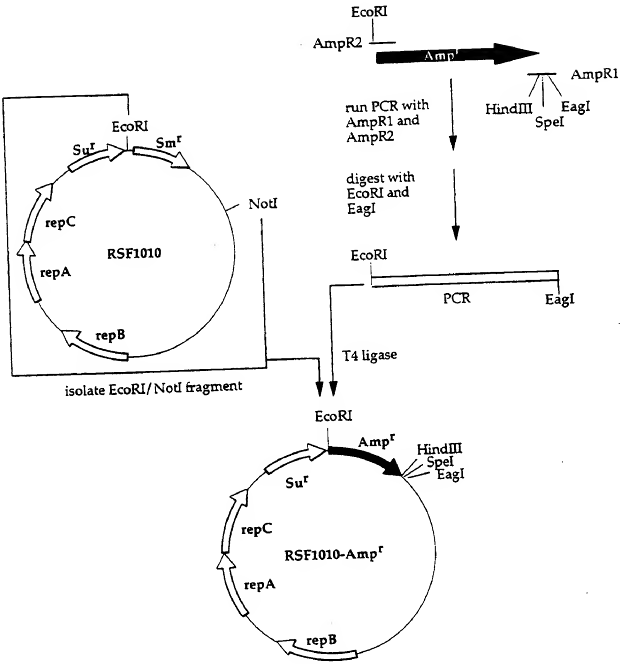
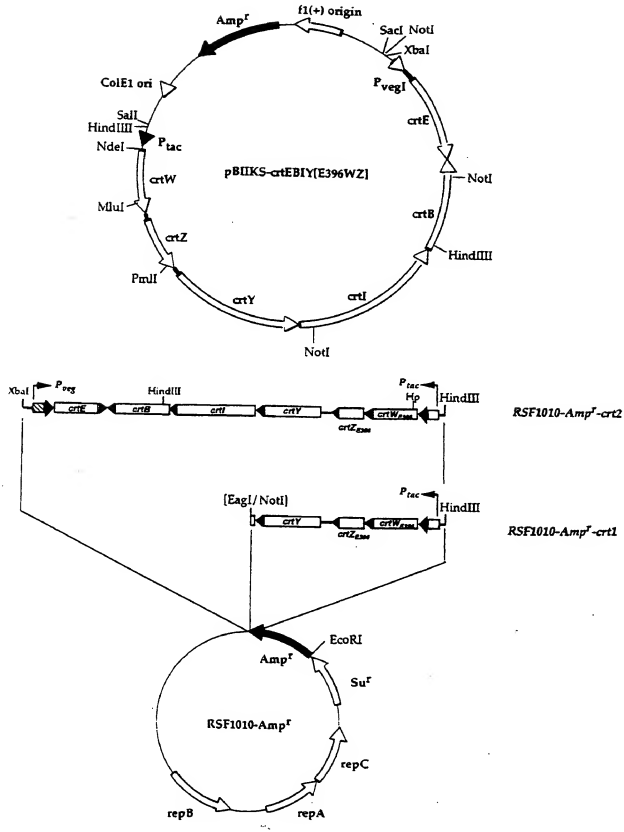

EcoRI
AmpR2
Amp^r
AmpR1
HindIII
SpeI
EagI
run PCR with AmpR1 and AmpR2
digest with EcoRI and EagI
EcoRI
PCR
EagI
T4 ligase
EcoRI
Su^r
Sm^r
NotI
repC
RSF1010
repA
repB
isolate EcoRI/NotI fragment
EcoRI
Amp^r
HindIII
SpeI
EagI
Su^r
repC
RSF1010-Amp^r
repA
repB

Fig. 13


Amp^r
f1(+) origin
SacI NotI
XbaI
ColE1 ori
P vegI
SalI
HindIII
Ptac
NdeI
crtE
pBIIKS-crtEBIY[E396WZ]
crtW
NotI
MluI
crtB
crtZ
HindIII
PmlI
crtY
crtI
NotI
XbaI
Pveg
HindIII
crtE
crtB
crtI
crtY
Ptac
Hp
HindIII
crtW
crtZ
RSF1010-Amp^r-crt2
[EagI/NotI]
Ptac
HindIII
crtY
crtW
crtZ
RSF1010-Amp^r-crt1
EcoRI
Amp^r
Su^r
RSF1010-Amp^r
repC
repB
repA

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 872 554 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
07.06.2000 Bulletin 2000/23

(43) Date of publication A2:
21.10.1998 Bulletin 1998/43

(21) Application number: 97120324.5

(22) Date of filing: 20.11.1997

(51) Int. Cl.[7]: **C12N 15/52**, C12P 23/00, A23K 1/16, A23L 1/275, C12N 1/21

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **02.12.1996 EP 96810839**

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)

(72) Inventors:
- **Pasamontes, Luis**
  4632 Trimbach (CH)
- **Tsygankov, Yuri,**
  State Scient. Cnt. of Russian Fed
  Moscow 113545 (RU)

(74) Representative: **Braun, Axel et al**
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

(54) Improved fermentative carotenoid production

(57) The present invention is directed to processes for the preparation of canthaxanthin, adonixanthin, astaxanthin, a mixture of adonixanthin and astaxanthin and zeaxanthin by a cell which has been transformed by DNA sequences encoding the respective biosynthetic enzymes of Flavobacterium and the gram negative bacterium E-396. Furthermore the present invention is directed to a food or feed composition comprising one or more of the aforementioned carotenoids.

Printed by Xerox (UK) Business Services
2.16.7 (HRS)/3.6

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 97 12 0324

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
D,A	EP 0 735 137 A (KIRIN BREWERY ;MARINE BIOTECH INST CO LTD (JP)) 2 October 1996 (1996-10-02) * page 5, line 55 - page 7, line 32 * * page 10 - page 11 * * page 14 - page 21; examples 7,8,12,13 *	1-7	C12N15/52 C12P23/00 A23K1/16 A23L1/275 C12N1/21
A	EP 0 635 576 A (NIPPON OIL CO LTD) 25 January 1995 (1995-01-25) * examples *	1-7	
D,A	WO 91 13078 A (AMOCO CORP) 5 September 1991 (1991-09-05) * page 24 - page 43 * * page 55 - page 64 * * page 164 - page 176; examples 21-24 *	3,5-7	
P,X	EP 0 747 483 A (HOFFMANN LA ROCHE) 11 December 1996 (1996-12-11) * the whole document *	2,3,5-7	TECHNICAL FIELDS SEARCHED (Int.Cl.6)
P,X	PASAMONTES L ET AL: "Isolation and characterization of the carotenoid biosynthesis genes of Flavobacterium sp. strain R1534" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, vol. 185, no. 1, 31 January 1997 (1997-01-31), pages 35-41, XP004093151 ISSN: 0378-1119 * the whole document *	3,6	C12N C12P

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	19 April 2000	Andres, S

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 97 12 0324

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

19-04-2000

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0735137 A	02-10-1996	AU 702584 B	25-02-1999
		AU 1281195 A	17-07-1995
		FI 962633 A	23-08-1996
		NO 962689 A	27-08-1996
		US 5811273 A	22-09-1998
		CA 2180024 A	06-07-1995
		WO 9518220 A	06-07-1995
		US 5972690 A	26-10-1999
EP 0635576 A	25-01-1995	CA 2128549 A	23-01-1995
		JP 7079796 A	28-03-1995
		NO 942731 A	23-01-1995
		US 5607839 A	04-03-1997
		US 5858761 A	12-01-1999
WO 9113078 A	05-09-1991	CA 2055447 A	03-09-1991
		EP 0471056 A	19-02-1992
		JP 5504686 T	22-07-1993
		US 5545816 A	13-08-1996
		US 5530188 A	25-06-1996
		US 5530189 A	25-06-1996
		US 5684238 A	04-11-1997
		US 5618988 A	08-04-1997
		US 5656472 A	12-08-1997
EP 0747483 A	11-12-1996	JP 9023888 A	28-01-1997

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82